

MARYUALÊ MALVESSI MITTMANN

**Construções de alçamento a sujeito:  
variação e gramaticalização**

**FLORIANÓPOLIS**  
**Março, 2006**

MARYUALÊ MALVESSI MITTMANN

## **Construções de alçamento a sujeito: variação e gramaticalização**

Dissertação apresentada ao Programa de Pós-Graduação em Lingüística da Universidade Federal de Santa Catarina como requisito parcial para obtenção do grau de Mestre em Lingüística.

Orientadora: Profª Drª Edair M. Gorski

**FLORIANÓPOLIS**  
**Março, 2006**



### *Agradecimentos*

À Universidade Federal de Santa Catarina e ao Programa de Pós-Graduação em Lingüística, na pessoa dos colegas, professores e funcionários.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

Aos membros da banca, Prof. Dr. Marco Rocha e Prof<sup>a</sup> Dr<sup>a</sup> Ana Zilles.

Aos colegas do Projeto VARSUL com quem mais convivi nestes últimos anos: Raquel, Priscilla, Guilherme, Gésyka, Marineide, Joana e Marco Antônio.

À minha orientadora Prof<sup>a</sup> Edair Gorski, pela confiança e competência, um agradecimento especial. Muitos são professores, poucos são exemplos.

Ao Fabio e à Gerciana, pela amizade e apoio “técnico”, e à Katja e ao Marco Aurélio, pela amizade e carinho em horas necessárias.

Ao Rodrigo, que nunca ficou realmente distante.

À minha família, por tudo.

*Muito obrigada.*

Só o desejo inquieto, que não passa,  
Faz o encanto da coisa desejada.  
E terminamos desdenhando a caça  
Pela doida aventura da caçada.

Mario Quintana

## RESUMO

Investigam-se sentenças complexas do português do Brasil, constituídas de uma oração matriz com verbo estativo na terceira pessoa do singular seguido de adjetivo, combinada com uma oração composta por verbo transitivo em forma reduzida e pelo complemento deste verbo (ex.: “É difícil conhecer o Brasil”). A principal característica destas construções é o fato do complemento do verbo infinitivo poder ser anteposto ao verbo finito, caracterizando então uma construção de alçamento a sujeito, bem como a possibilidade de uso de diferentes formas de conexão entre estas orações: justaposição, preposição *de* ou preposição *para*, (exs.: O Brasil é difícil [de/para] conhecer). O propósito desta pesquisa é estudar a variação no uso das diferentes configurações dessas sentenças complexas, bem como a variação nas formas de conexão entre a oração matriz e a combinada, e verificar se existe alguma relação entre configurações e formas de conexão específicas. Esta pesquisa é norteada pela perspectiva da Sociolinguística Variacionista, tomando-se como teoria de base o funcionalismo de Givón e o paradigma da gramaticalização como motriz da fixação de estruturas gramaticais. Os dados utilizados foram coletados em corpus de fala da Região Sul do Brasil (extraído do banco de dados do Projeto VARSUL) e na escrita padrão de jornais de prestígio nacional (extraído do banco de dados do NILC). As principais hipóteses são de que diferentes configurações estão principalmente relacionadas a diferentes graus de topicalidade do complemento do verbo infinitivo e que a forma de conexão escolhida reflete diferentes graus de integração entre as orações. Levantam-se diversos fatores de natureza linguística e extra-linguística que podem exercer influência sobre a ordem dos constituintes das sentenças complexas em estudo, bem como sobre a variação nas formas de conexão entre as orações. Os métodos utilizados para esta investigação se baseiam na Sociolinguística Variacionista e na Linguística de Corpus. Os dados foram coletados com auxílio do pacote WordSmith Tools, em todos os textos que compõem os corpora selecionados para esta pesquisa. Foram aplicados testes de significância e associação no cruzamento da variável dependente com cada uma das variáveis independentes, além de realizado o cálculo dos pesos relativos de cada fator em função da aplicação da regra de diferentes configurações e formas de conexão. Os resultados indicam que a ocorrência de construções alçadas é influenciada pelo uso do conector *de*, como exemplificado por “O Brasil é difícil de conhecer”, construções cuja estrutura revela alto grau de integração.

**Palavras chave:** alçamento a sujeito, conectores oracionais, integração de orações.

## ABSTRACT

This research deals with complex sentences in spoken and written Brazilian Portuguese. The sentences analyzed are composed by a finite main clause with a stative verb followed by an adjective, combined with a non-finite clause that consists of a transitive verb and its complement (ex: “*É difícil conhecer o Brasil*” / It is difficult to know Brazil). The most important feature of these kind of construction is the fact that the verbal complement of the non-finite clause can be placed in the first position of the main clause, which characterizes a raising-to-subject construction. Another important factor is the possibility (in Portuguese) of using different connectors in order to combine the two clauses: juxtaposition, the preposition “*de*” and the preposition “*para*” (exs. *O Brasil é difícil [de/para] conhecer* / Brazil is difficult to know). The aim is to study the variation in the use of different syntactic configuration in these complex sentences, as well as the variation in the forms of connection of the main and the embedded clauses, and to investigate if there is any relation between a particular syntactic configuration and one of the forms of connection. This study is oriented by the Variacionist Sociolinguistics, having as basis theory the functionalism of Givón, and also the grammaticalization paradigm as the background to the stabilization of grammatic structures. The data was collected in two different corpora: one of the southern Brazil speech (extracted from Projeto VARSUL database) and another of Brazilian prestigious newspapers (extracted from NILC database). The main hypothesis are that different configurations are related to different degrees of the topicality of the non-finite verb complement, and that the connection form selected reflects different degrees of clause integration. A variety of linguistic and extra-linguistic factors were considered, that may influence on the studied sentences word order, and also on the clause connection forms variation. The methods employed here are based in Variacionist Sociolinguistics and in the Corpus Linguistics. The data were collected with the software *WordSmith Tools*, in all the texts selected to form the corpora to this investigation. Significance and association tests were applied, and the relative weights for different syntactic configurations and connection forms was also calculated. The results point that the use of raised constructions is related to the preposition “*de*”, such as in the example: “*O Brasil é difícil de conhecer*” (Brazil is difficult to know), which structure reveals high degree of clause integration.

**Keywords:** raising to subject, clause connectors, clause integration.

## SUMÁRIO

	P.
LISTAS DE FIGURAS.....	9
APRESENTAÇÃO.....	11
1 O FENÔMENO EM ESTUDO.....	13
1.1 Delimitação do objeto.....	15
2 REVISÃO DA LITERATURA.....	17
2.1 A visão tradicional.....	18
2.2 Variação e gramaticalização em uma perspectiva funcional.....	21
2.3 Combinação de orações.....	27
2.4 Topicalidade e alçamento do SN.....	30
2.5 Estudos Preliminares.....	32
3 PROCEDIMENTOS METODOLÓGICOS.....	37
3.1 Objetivos.....	37
3.2 Questões e hipóteses.....	38
3.3 Materiais e Métodos.....	40
3.3.1 O banco de dados de fala.....	41
3.3.2 O banco de dados de escrita.....	42
3.3.3 Coleta de dados.....	44
3.3.4 Variáveis linguísticas.....	46
3.3.5 Variáveis sociais.....	55
3.3.6 Tratamento estatístico.....	56
4 RESULTADOS E DISCUSSÃO.....	59
4.1 Resultados da amostra de textos escritos (NILC).....	66
4.1.1 Configuração da sentença em textos escritos.....	67
4.1.1.1 Alçamento e forma de conexão entre as orações na escrita.....	68
4.1.1.2 Alçamento e forma do SN na escrita.....	69
4.1.1.3 Ordem canônica: forma do SN e expressão do sujeito na escrita.....	71
4.1.1.4 Distância referencial e status informacional na escrita.....	72
4.1.2 Forma de conexão em textos escritos (NILC).....	73
4.1.2.1 Forma de conexão e forma do SN na escrita.....	74
4.1.2.3 Forma do item qualificador na escrita.....	75
4.1.2.4 Forma de conexão e Expressão do sujeito do infinitivo na escrita.....	76
4.2 Resultados da amostra de fala da Região Sul (VAR SUL).....	77
4.2.1 Configuração da sentença na fala da Região Sul (VAR SUL).....	78
4.2.1.1 Alçamento e forma de conexão entre as orações na fala.....	80
4.2.1.2 Forma do SN na fala.....	81
4.2.1.3 Alçamento e expressão do sujeito do infinitivo na fala.....	82
4.2.1.4 Distância referencial e status informacional na fala.....	83
4.2.1.5 Ordem canônica e forma de conexão na fala.....	85
4.2.1.6 Ordem canônica e expressão de sujeito do infinitivo.....	85
4.2.1.7 Ordem canônica e status informacional.....	86
4.2.1.8 Ordem canônica e escolaridade na fala.....	87
4.2.1.9 Deslocamento da oração infinitiva na fala.....	88
4.2.2 Forma de conexão entre as orações na fala da Região Sul (VAR SUL).....	89
4.2.2.1 Preposição <i>de</i> na fala.....	91
4.2.2.2 Preposição <i>para</i> na fala.....	92
4.2.2.3 Justaposição na fala.....	93
4.3 Resultados da amostra de fala de Florianópolis (VAR SUL).....	95
5 CONSIDERAÇÕES FINAIS.....	100
REFERÊNCIAS.....	106



## LISTAS DE FIGURAS

<b>DIAGRAMAS</b>	<b>p.</b>
Diagrama 1.....	102
Diagrama 2.....	104
<b>GRÁFICOS</b>	
Gráfico 1	
Configuração da sentença na amostra de fala (VARSUL).....	65
Gráfico 2	
Configuração da sentença na amostra de escrita (NILC).....	65
<b>TABELAS</b>	
Tabela 1	
Distribuição do tipo de configuração da sentença nas amostras sincrônica e diacrônica.....	34
Tabela 2	
Distribuição da forma de conexão entre as orações na amostra diacrônica.....	35
Tabela 3	
Distribuição da forma de conexão entre as orações na amostra sincrônica por cidade.....	35
Tabela 4	
Variável configuração da sentença em toda a amostra (NILC e VARSUL).....	61
Tabela 5	
Variável configuração da sentença de acordo com a origem das ocorrências (NILC ou VARSUL).....	64
Tabela 6	
Variável configuração da sentença na amostra de escrita (NILC).....	67
Tabela 7	
Alçamento vs. Forma de conexão – pesos relativos da amostra de escrita.....	69
Tabela 8	
Alçamento vs. Forma do SN – pesos relativos da amostra de escrita.....	70
Tabela 9	
Ordem canônica vs. Forma de conexão – pesos relativos da amostra de escrita.....	71
Tabela 10	
Ordem canônica vs. Expressão do sujeito do inf. – pesos relativos da amostra de escrita.....	71
Tabela 11	
Variável forma de conexão entre as orações na amostra de escrita (NILC).....	73
Tabela 12	
Preposição <i>de</i> vs. Forma do SN – pesos relativos da amostra de escrita.....	74
Tabela 13	
Preposição <i>de</i> vs. Forma do item qualificador – pesos relativos da amostra de escrita.....	75
Tabela 14	
Preposição <i>para</i> vs. Expressão do sujeito do inf. – pesos relativos da amostra de escrita.....	76
Tabela 15	
Variável configuração da sentença na amostra de fala da Região Sul – fatores sociais.....	78
Tabela 16	
Variável configuração da sentença na amostra de fala da Região Sul – fatores lingüísticos.....	79
Tabela 17	
Alçamento vs. Forma de conexão – pesos relativos da amostra de fala.....	80
Tabela 18	
Alçamento vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala.....	82
Tabela 19	
Ordem canônica vs. Forma de conexão – pesos relativos da amostra de fala.....	85
Tabela 20	
Ordem canônica vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala.....	85
Tabela 21	
Ordem canônica vs. Status informacional – pesos relativos da amostra de fala.....	86
Tabela 22	
Ordem canônica vs. Nível de escolaridade – pesos relativos da amostra de fala.....	87
Tabela 23	
Oração infinitiva deslocada vs. Forma de conexão – pesos relativos da amostra de fala.....	88

Tabela 24	
Oração infinitiva deslocada vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala.....	89
Tabela 25	
Variável forma de conexão entre as orações na amostra de fala da Região Sul – fatores sociais.....	89
Tabela 26	
Variável forma de conexão entre as orações na amostra de fala da Região Sul – fatores lingüísticos.....	90
Tabela 27	
Preposição <i>de</i> vs. Forma do SN – pesos relativos da amostra de fala.....	92
Tabela 28	
Preposição <i>para</i> vs. Expressão do sujeito do inf. – pesos relativos da amostra de fala.....	93
Tabela 29	
Justaposição vs. Forma do SN – pesos relativos da amostra de fala.....	94
Tabela 30	
Justaposição vs. Pres. de materiais intervenientes – pesos relativos da amostra de fala.....	94
Tabela 31	
Variável configuração da sentença na amostra de fala de Florianópolis – fatores sociais.....	96
Tabela 32	
Variável configuração da sentença na amostra de fala de Florianópolis – fatores lingüísticos.....	97
Tabela 33	
Variável forma de conexão entre as orações na amostra de fala de Florianópolis.....	98

## QUADROS

Quadro 1	
Esquema da configuração das sentenças combinadas em estudo.....	16
Quadro 2	
Características das amostras de língua falada e escrita.....	41
Quadro 3	
Distribuição dos informantes em células por cidade, faixa etária sexo e escolaridade.....	55
Quadro 4	
Distribuição dos informantes de Florianópolis em células por cidade, faixa etária sexo e escolaridade.....	56
Quadro 5	
Testes de significância e associação – Configuração da sentença vs. Forma de conexão (escrita).....	68
Quadro 6	
Testes de significância e associação – Configuração da sentença vs. Forma do SN (escrita).....	70
Quadro 7	
Testes de significância e associação – Config. da sentença vs. Distância referencial (escrita).....	72
Quadro 8	
Testes de significância e associação – Config. da sentença vs. Status informacional (escrita).....	72
Quadro 9	
Testes de significância e associação – Forma de conexão vs. Forma do SN (escrita).....	74
Quadro 10	
Testes de significância e associação – Forma de conexão vs. Forma do item qualificador (escrita).....	75
Quadro 11	
Testes de significância e associação – Forma de conexão vs. Exp. do sujeito do infinitivo (escrita).....	76
Quadro 12	
Testes de significância e associação – Configuração da sentença vs. Forma de conexão (fala).....	80
Quadro 13	
Testes de significância e associação – Configuração da sentença vs. Forma do SN (fala).....	81
Quadro 14	
Testes de significância e associação – Configuração da sentença vs. Exp. do sujeito do infinitivo (fala).....	82
Quadro 15	
Testes de significância e associação – Configuração da sentença vs. Distância referencial (fala).....	83
Quadro 16	
Testes de significância e associação – Configuração da sentença vs. Status informacional (fala).....	84
Quadro 17	
Testes de significância e associação – Forma de conexão vs. Forma do SN (fala).....	91
Quadro 18	
Testes de significância e associação – Forma de conexão vs. Presença de materiais intervenientes (fala).....	94

## APRESENTAÇÃO

O trabalho de pesquisa “Construções de alçamento a sujeito: variação e gramaticalização” está inserido na linha de pesquisa “Variação, Mudança e Ensino”, da área de concentração Sociolingüística, e está atrelado ao projeto da Profª Edair Gorski “Não está fácil ((pra/de) a gente) viver aqui: gramaticalização e variação de construções subjetivas”, ao qual estão vinculados os sub-projetos de pesquisa de iniciação científica de Mittmann (2003), Neves (2003; 2004), e May (2004). Estes estudos preliminares demonstraram a complexidade de um tipo particular de construções em português, que apresentam em sua configuração básica: verbo estativo conjugado na terceira pessoa do singular (ser, estar, ficar), seguido de um constituinte nominal (adjetivo ou substantivo) formando uma oração matriz, que encabeça uma oração combinada formada por um verbo no infinitivo (pessoal ou impessoal) seguido de complemento (ex.: “É difícil conhecer o Brasil”). Outros elementos também podem estar presentes, tais como advérbios (Não é difícil conhecer o Brasil / É muito difícil conhecer o Brasil), sujeito do verbo infinitivo (É difícil a gente conhecer o Brasil), e principalmente, os conectores *de* e *para* (É difícil de/para conhecer o Brasil).

Os trabalhos já realizados também apontam para diferenças no uso dessas construções nas capitais da Região Sul (Curitiba, Florianópolis e Porto Alegre), e além disso, verificam o uso de tais sentenças complexas no português europeu antigo. Pela complexidade do funcionamento desse tipo de construção e suas variações, sentiu-se a necessidade de aprofundar os estudos realizados até então, delimitando com mais precisão o objeto a ser investigado, ampliando-se as fontes de pesquisa e analisando mais a fundo os fatores gramaticais, discursivos e sociais que podem estar atuando nas diferenças encontradas no uso dessas estruturas.

Uma das características que chamou a atenção para este tipo de construção durante as pesquisas anteriores é que, em alguns casos, notou-se a ocorrência da mesma construção básica, mas com configurações sintáticas diferentes. Algumas vezes, a oração do verbo

infinitivo é anteposta à matriz (ex.: Conhecer o Brasil é difícil), e em outros casos, era o complemento do verbo infinitivo que estava deslocado para o início da sentença (ex.: O Brasil é difícil de conhecer). De modo a delimitar o objeto da pesquisa, o foco do presente trabalho recai apenas nas construções em que se verificou a possibilidade de uso de qualquer destas configurações. Assim, restringiu-se a busca nos dados para as sentenças com a estrutura apresentada, mas que apresentassem verbo infinitivo transitivo, acompanhado de complemento, ou cujo complemento estivesse facilmente acessível nas orações precedentes.

Isso abriu uma nova possibilidade de análise, que é a de investigar a relação de aspectos funcionais da organização do discurso<sup>1</sup>, tanto oral como escrito, com a organização sintática de uma sentença complexa. Questiona-se em que medida diferentes estratégias de organização do discurso, no caso específico, a distribuição do referente codificado como complemento do verbo infinitivo ao longo de um trecho de discurso (nos exemplos o elemento “o Brasil”) pode influenciar a fixação de uma estrutura particular, através de um processo de gramaticalização. Como se trata de um objeto ainda pouco explorado, também buscou-se verificar se há outros fatores que possam influenciar a escolha de uma configuração sintática particular, tanto sociais como lingüísticos, gerando assim novas questões para serem aprofundadas em pesquisas futuras.

Deste modo, esta dissertação alinha-se às pesquisas sobre variação e gramaticalização de sentenças complexas e visa contribuir com os estudos gramaticais que contemplam especialmente a questão da ordem dos constituintes oracionais e do uso de conectores em orações combinadas. Num sentido mais amplo, busca-se contribuir com a descrição do português falado nas capitais da Região Sul, bem como da modalidade escrita padrão da língua atualmente empregada no Brasil, assim como ampliar as possibilidades de pesquisa na área.

---

1 O termo “discurso” é utilizado neste trabalho referindo-se ao uso concreto da língua, seja na modalidade oral ou escrita. Procura-se com esta terminologia abarcar os enunciados que superam os limites do chamado período composto, já que se parte da idéia de que existem mecanismos reguladores da organização dos enunciados, qualquer que seja a sua extensão. Estes mecanismos são tanto de ordem funcional e pragmática, na medida em que o locutor organiza seu enunciado levando em consideração o interlocutor; quanto cognitiva, já que o processamento verbal é subordinado aos limites de capacidades mentais tais como memória de trabalho e atenção.

## 1 O FENÔMENO EM ESTUDO

Quando se observa o uso de sentenças complexas<sup>2</sup>, o pesquisador se depara, muitas vezes, com construções que, se por um lado são facilmente interpretáveis do ponto de vista semântico-pragmático, por outro lado, apresentam maiores dificuldades na hora de se realizar uma análise do ponto de vista sintático, ao buscar identificar as funções dos constituintes, ou avaliar o tipo de relação existente entre as orações envolvidas, se de relativa autonomia ou de interdependência. Dentro do amplo espectro de sentenças complexas em língua portuguesa, destacam-se aqui aquelas ilustradas nos exemplos<sup>3</sup> abaixo:

- (1) É difícil ele comprar uma cerveja. (VARSQL, FLP J21)<sup>4</sup>
- (2) Fórmula um hoje é difícil ver, né? (VARSQL, FLP J2)
- (3) Olha, churrasco é bem fácil de preparar, né? (VARSQL, POA 04)
- (4) Porque hoje, pra ter uma amiga, não está fácil, não (VARSQL, FLP 07).

As dúvidas na análise de sentenças como estas surge, principalmente, quando o analista se depara com as diferentes configurações possíveis para estas sentenças, e também com a preposição precedendo o verbo infinitivo, como nos exemplos (3) e (4). Nestes casos, a preposição está funcionando como conector entre as orações? É possível aplicar a mesma análise a todas as sentenças? Que implicações a preposição empregada acarreta na análise e interpretação do enunciado? As mesmas sentenças poderiam ser enunciadas sem a preposição e sem alterações no significado? Até que ponto é possível mexer na ordenação dos constituintes sem alterar o significado?

Retomando os enunciados acima, é possível perceber mais claramente a ambigüidade existente nestas sentenças, pois é difícil estabelecer com certeza qual a vinculação existente

2 A expressão “sentença” é utilizada aqui no mesmo sentido em que são empregados “enunciado” e “construção”, referindo-se sempre à oração matriz acompanhada da oração combinada (por este motivo complexa). O termo “oração” é empregado em sua acepção tradicional.

3 Exemplos extraídos do banco de dados VARSQL (ver seção 4 “Materiais e Métodos”).

4 O código entre parênteses indica a cidade e o número do informante da entrevista onde se encontra a ocorrência.

entre as orações. Ao reescrevê-las, percebe-se que a presença do conector não altera o sentido dos enunciados.

(1) É difícil ele **comprar uma cerveja**.

- a. [É difícil [ele comprar [uma cerveja]]]
- b. [É difícil pra ele [comprar uma cerveja]]
- c. [[Pra ele comprar uma cerveja] [é difícil]]

(2) **Fórmula um** hoje é difícil **ver**, né?

- a. [Fórmula um hoje é difícil [ver]]
- b. [Hoje é difícil [ver fórmula um]]
- c. [[Fórmula um] [hoje é difícil de ver]]

(3) Olha, **churrasco** é bem fácil **de preparar**, né?

- a. [Churrasco é bem fácil [de preparar]]
- b. [É bem fácil [de preparar churrasco]]
- c. [[Churrasco,] [é bem fácil [preparar t,]]]

(4) Porque hoje, **pra ter uma amiga**, não está fácil, não.

- a. [[Pra ter uma amiga] não está fácil]
- b. [[Pra ter] [uma amiga não está fácil]]
- c. [ter [uma amiga não está fácil]]

O que se tenta mostrar através da reestruturação das sentenças é que existe mais de uma interpretação possível para a hierarquia dos constituintes. Em alguns casos, acredita-se que esta pluralidade seja reflexo de um fenômeno de variação, desencadeado por um processo de reanálise estrutural das sentenças. Acredita-se, então, que sentenças complexas como as exemplificadas acima não são formas totalmente independentes no sistema lingüístico, mas que seu funcionamento e uso pode ser compreendido dentro de uma regra variável. Se este for o caso, a introdução de um conector serviria para explicitar a relação estrutural dos constituintes das orações. Considerando-se este um fenômeno de variação, espera-se que cada forma de conexão assuma um contexto mais específico de uso.

Como é possível observar nos exemplos acima, em muitos casos é possível formular

estes enunciados com ou sem a presença do conector, sem que isso acarrete mudança do significado proposicional dos enunciados. Isto ocorre mesmo quando é introduzida a preposição *para*, a qual ainda apresenta uma maior delimitação de sentido, se comparada à preposição *de*, cujo uso é muito mais generalizado, já tendo sofrido maior desgaste semântico. Em casos como (3c), alguns falantes estranham a construção sem a preposição *de*, ainda que este tipo de sentença (sem a preposição) tenha sido encontrado no corpus do VARSUL (KNIES, COSTA, 1995), na cidade de Florianópolis<sup>5</sup>.

Do ponto de vista da função comunicativa, ou da significação, os enunciados parecem se caracterizar em geral por trazer, na oração finita, uma avaliação ou apreciação particular sobre a situação descrita na oração infinitiva, cujo conteúdo é factual. O enunciador constrói uma asserção categórica acerca daquilo que está sendo avaliado, o que ocorre na grande maioria das vezes sem modalização que atenua o que está sendo enunciado. Isso sugere que o enunciador tem um grande compromisso e envolvimento emocional com aquilo que está atestando, o que pode fazer estas expressões mais passíveis de serem sujeitas a princípios pragmáticos de organização do discurso.

### 1.1 Delimitação do objeto

Propõe-se realizar um estudo sobre sentenças complexas do português, cuja característica principal é a possibilidade de, em tais sentenças, ocorrer o alçamento<sup>6</sup> (raising) do constituinte que ocupa a função de objeto direto na oração combinada para a posição do sujeito da oração matriz, como em (5) e (5a) abaixo.

(5) É difícil resolver **este problema**.

a. **Este problema** é difícil resolver.

As sentenças devem possuir em sua configuração os elementos descritos a seguir:

5 Maiores detalhes na seção 1.2 “Estudos preliminares”.

6 Na seção 2.3 “Topicalidade e alçamento do SN” justifica-se a opção por tratar ambas as formas como variantes.

**a) Oração matriz:**

Contém um verbo estativo (tipicamente *ser*, *estar* e *ficar*) na terceira pessoa do singular, seguido de um constituinte nominal, que será aqui denominado ‘item qualificador’ (IQ) da oração combinada (geralmente um adjetivo).

**b) Oração combinada:**

Apresenta verbo transitivo na forma infinitiva, o qual pode ou não apresentar um sujeito expreso. O complemento (objeto) do verbo deve estar expreso, o que ocorre geralmente na forma de um sintagma nominal (SN). Quando não estiver expreso na sentença, o complemento deve ser facilmente recuperado no contexto, ou seja, deve ter sido mencionado no discurso precedente.

**c) Conexão entre as orações:**

Pode se dar através da simples justaposição (ausência de conector) ou através de um dos conectores, a preposição *de* ou a preposição *para*<sup>7</sup>.

O quadro 1 apresenta um esquema para a configuração canônica<sup>8</sup> destas sentenças. Os constituintes entre parênteses não são obrigatórios, e o complemento verbal (objeto), representado entre colchetes, é o elemento que pode sofrer o alçamento.

**Quadro 1**

**Esquema da configuração das sentenças combinadas em estudo**

**VP3+IQ+(PREP)+(SUJ)+INF+[COMP]**

Ex.: É fácil (de/pra) (a gente) fazer [café].

Duas premissas norteiam o desenvolvimento desta pesquisa. Uma é que as sentenças com diferentes formas de conexão podem ser consideradas variantes, do mesmo modo como a forma na qual o sintagma nominal se encontra à esquerda do verbo finito pode ser considerada variante da forma que apresenta o sintagma nominal à direita do verbo infinitivo, a outra segunda é que esses dois fenômenos de variação estão correlacionados, e podem ser compreendidos dentro de um processo de gramaticalização, como será justificado adiante.

<sup>7</sup> Em casos muito raros foi documentada a ocorrência da preposição *a* no português antigo e do século XVIII (MITTMANN, 2003; GORSKI, MITTMANN, MAY, 2004). Não foi encontrado registro deste conector nos textos contemporâneos pesquisados, tanto falados quanto escritos.

<sup>8</sup> Considera-se a ordem verbo estativo seguido de adjetivo seguido do verbo infinitivo seguido do complemento a ordem canônica desta sentença por ser aquela que aparece no maior número de ocorrências, tanto nos levantamentos diacrônicos quanto sincrônicos.



## 2 REVISÃO DA LITERATURA

A primeira parte desta revisão irá reunir as observações de gramáticas tradicionais sobre as sentenças complexas em estudo (seção 2.1), as seguintes (seções 2.2 a 2.3) são dedicadas ao referencial teórico utilizado como suporte da análise e método de investigação realizados aqui. A última seção deste capítulo (2.5) destina-se a resumir os resultados obtidos nos estudos realizados anteriormente.

Procura-se sustentar este trabalho a partir dos referenciais teóricos da sociolingüística variacionista e da gramática funcional. Na primeira, busca-se apoio na investigação da variação no uso das construções de alçamento, bem como da forma de conexão entre as orações. A segunda, por sua vez, se faz presente por ser a visão de gramática que mais parece dar conta da tarefa de explicar fenômenos de natureza discursiva, bem como sustentar o conceito de "gramaticalização" utilizado aqui. Além disso, ambas as abordagens procuram enfatizar o uso e a frequência de ocorrência das formas lingüísticas, formando portanto um alinhavo teórico apropriado para a proposta apresentada nesta dissertação.

Através deste panorama teórico, busca-se compreender a variação na codificação das sentenças reduzidas analisadas neste trabalho, bem como as relações gramaticais envolvidas na composição destas estruturas. As variações observadas nestas sentenças (ordem dos constituintes oracionais e formas de conexão entre as orações) são consideradas fenômenos relacionados entre si, devendo ser compreendidos a partir de uma teoria da linguagem abrangente. Parte-se do princípio de que a variação surge de uma necessidade pragmático-discursiva no que se refere às questões de ordem, como também das próprias características do sistema gramatical do português, no que se refere à introdução de itens gramaticais como as preposições.

## 2.1 A visão tradicional

Consultando manuais de gramática em busca de algum tipo de classificação ou descrição para as estruturas aqui estudadas, o mais próximo que se chega ao tipo de sentença aqui apresentada é a descrição de estruturas subordinadas subjetivas. Oiticica (1940, p. 201) apresenta alguns “esquemas de construção” para as orações subordinadas substantivas subjetivas, dentre os quais o “esquema” proposto em a abaixo é o que mais se aproxima da estrutura encontrada nas sentenças em discussão. Descrição similar também é referida em Rocha Lima (1986, p. 235):

- a) Verbo da oração principal (ser, estar, ficar) em terceira pessoa seguido de substantivo ou adjetivo;
- b) Verbo com sentido de conveniência (convém), dúvida (parece), ocorrência (acontece) e efeito moral (agrada);
- c) Verbo na voz passiva (pronominal ou analítica).

Sentenças complexas com orações infinitivas como as exemplificadas anteriormente são geralmente classificadas nas gramáticas tradicionais como um período composto por subordinação, em que a oração infinitiva é denominada *oração substantiva subjetiva reduzida de infinitivo* (ALMEIDA, 1962; ROCHA LIMA, 1972; CUNHA, 1977). A classificação enquanto subjetiva, no entanto não é um consenso. Para Sacconi (1985, p. 291), infinitivos ligados a adjetivos são considerados complementos nominais, e segundo o autor, neste caso não há oração reduzida, exemplificando seu ponto de vista com construções como “roupa fácil de passar”, “livro difícil de ler”. O autor não considera a hipótese de construções como “é difícil de passar roupa”, ou “é difícil de ler o livro”, e não menciona orações reduzidas subjetivas.

O uso da preposição antes do verbo infinitivo constitui um ponto de bastante discussão dentro das gramáticas tradicionais. Bechara (1976, p. 237) observa: "Por vezes a oração reduzida subjetiva ou objetiva direta tem o seu infinitivo precedido de preposição expletiva: 'Desaire real seria de a deixar sem prêmio', 'Custou-lhe muito a aceitar a casa'" (grifos meus). A respeito deste exemplo, Melo (1978, 218-219) comenta que a preposição aparece por analogia (ou cruzamento sintático) com frases em que o infinitivo

preposicionado vem completando o substantivo: “teve desejo de fazer”. Segundo o autor, a circunstância de o infinitivo-sujeito vir posposto dá a **impressão psicológica** de que ele é **objeto direto**, daí poder vir preposicionado (grifos meus).

Para outros gramáticos tais como Epiphanyo Dias (1970), Said Ali (1971), Rocha Lima (1972) e Cunha e Cintra (1985), o infinitivo precedido de preposição adquire um “sentido passivo”.

Em Dias (1970, p. 228) encontramos:

Emprega-se qualificativamente o infinitivo precedido da prepos. de:  
[...] Aos adjetivos facil, difficil, bem máo, duro, aspero, raro e outros de significação semelhante, e a longo, gostoso liga-se o simples infinitivo precedido da prepos. de em sentido limitativo (v. § 169) e, aparentemente, passivo.

Said Ali (1971, p.341-2) comenta:

Combinado com certos adjetivos ou com o verbo ser ou estar, o infinitivo preposicionado pode ter sentido passivo, correspondendo umas vezes ao supino -u, outras ao gerúndio latino:  
Isto procedia de ser elle mau de contentar (Barros).  
As cousas não eram para dizer nem são para ouvir (Vieira).  
Esta mesma representação he muyto difficultosa de conservar (Vieira) Isto que tão facilmente se diz, não he tão facil de entender (Vieira). Esta virtude he mais facil de louvar que de descobrir (Freire).

Em obra anterior o autor detalha os usos os usos do infinitivo com valor passivo em construções do seguinte tipo (ALI 1965, p. 174):

Depois dos dizeres fácil de, difícil de, bom de, mau de, impossível de, digno de e alguns outros adjetivos seguidos da preposição de:  
Êste trecho latino é difícil de entender e pior de traduzir.  
Emudeçam arbítrios e discursos fáceis de escrever e impossíveis de executar (Vieira).  
Obs.: Com o adjetivo digno, o verbo também pode aparecer na forma passiva ou reflexiva: Daqui se inferem duas conseqüências muito dignas de ser notadas (Vieira).  
depois das expressões estar por, ficar por, para designar ação ainda não efetuada:  
O prédio está por acabar.  
Ainda nos resta por vencer a maior dificuldade (Vieira)  
depois das expressões é de, é para, significando aquela “é cousa merecedora de”, ou “digna de”, e esta “é cousa própria para”:  
Não é de espantar lerem os católicos pelos livros de gentio (Vieira) Seria muito para desejar (Vieira)

Mais recentemente, Melo (1978, p.186) descreve o uso do infinitivo com sentido passivo em construções da seguinte natureza:

- (a) o infinitivo é complemento de certos adjetivos, tais como “duro”, “fácil”, “difícil”:  
Ossos duros de roer (= de serem roídos)  
Lições difíceis de aprender (= de serem aprendidas)
- (b) O infinitivo, regido de de, equivale a um adjetivo com sufixo -vel, significa “digno de ser...”: São de notar e imitar exemplos como estes de lealdade e abnegação (= notáveis; imitáveis)
- (c) O infinitivo vem regido de a ou para e apresenta sentido passivo:

As dificuldades são para resolver e não para contemplar (= para serem resolvidas, serem contempladas)

(d) O infinitivo não vem preposicionado, mas, apesar da forma ativa, tem sentido passivo: O mesmo dia os viu batizar (= serem batizados) (MA, Esaú e Jacó, p.27).

Rocha Lima (1972, p. 381) ao descrever o emprego do infinitivo não flexionado, também registra o uso com sentido passivo, regido de preposição *de* como complemento de um adjetivo: “**As cadeiras antigas**, pesadas e maciças **eram difíceis de menear** (Rui Barbosa)”, “**Versos! São bons de ler**, mais nada; eu penso assim (M. de Assis)” (grifos meus).

Tratando do emprego da forma não flexionada do infinitivo, Cunha e Cintra (1985, p. 475) destacam o fato de essa forma verbal servir de complemento nominal a adjetivos como fácil, possível, bom, raro e outros semelhantes: “Já não transitam pelo correio **aquelas cartas de letra miudinha, impossíveis de ler, gratas de ler**, pois derramavas nelas uma intacta ternura... (C.D. de Andrade)” (grifo meu). Em nota de rodapé, os autores registram:

Este emprego da forma não flexionada deve ser incluído, como ensina Theodoro Henrique Maurer Jr., entre aqueles, nos quais o infinitivo depende “de um adjetivo, de um substantivo ou de um verbo em construções em que corresponde a um supino em -u ou a um infinitivo passivo da língua latina. Exemplos: As crianças são fáceis de contentar; Estas nozes são boas de abrir; Ficaram algumas peças por cortar; Tais resultados eram de prever”. (O infinitivo flexionado português; estudo histórico-descritivo. São Paulo, Companhia Editora Nacional/ USP, 1968, p.236).

Cunha e Cintra mencionam ainda, em continuidade à nota:

Para o ilustre filólogo paulista [Maurer], que neste ponto acompanha o pensamento de Nyrop e Ferdinand Brunot, o infinitivo **não apresenta em tais construções o sentido passivo**, que costumam atribuir-lhe os gramáticos, mas “um sentido geral, em que a idéia de voz ativa ou passiva se apaga”.

Este “sentido passivo” a que se referem os autores, provavelmente decorre da supressão do agente em tais construções. Nos exemplos apresentados pelos gramáticos, as construções apresentam verbo transitivo em que se atribui posição de destaque para o complemento verbal e omite-se o agente, estratégia recorrente nas construções em voz passiva.

A introdução de infinitivo por preposição também é um fator que chamou a atenção de nossos gramáticos históricos. Epiphânio Dias (DIAS 1970, p. 217-219) comenta em sua *Syntaxe histórica portuguesa* (grifos meus):

Um dos factos mais notaveis relativos á syntaxe do infinitivo, nas lingoas romanicas, é a construção de preposições com o infinitivo. [...] Em port. as preposições constroem-se não só com o simples infinitivo, senão até com or. infinitivas (v. g.: por tu saberes). A

**construção de preposições com o infinitivo tornou-se tão familiar**, que, em português, e em outras línguas românicas (v. M. Lübke, §340), **chegam a antepôr-se a infinitivos que exercitam as funções de sujeito, facto de que não se deu ainda explicação satisfatória**. Do emprego de preposições com o infinitivo no baixo latim cita Diez (Gr., Synt., IV, 8, 2) textos que ascendem às primeiras décadas do sec. VIII. [...] **No port. arch. medio este infinitivo é às vezes precedido da prep. de**.

De modo semelhante, Said Ali (ALI 1964, p. 339-340) também sentiu estranhamento ao encontrar tais construções (grifos meus):

Está ainda há espera de esclarecimento satisfatório o curioso fato, comum a outros idiomas modernos (ainda que em casos diversos do português), de antepor preposição a infinitivos dependentes de certos verbos, sendo entretanto tal o sentido destes termos regentes que parece dispensável a partícula. Em português usa-se antes do referido infinitivo ora a, ora de, podendo a sintaxe variar de uma época para outra, o que ainda mais complica a análise.

[...] Em português antigo podia-se calar entre dous verbos não somente a partícula inexpressiva, mas por vezes também a preposição claramente denotadora de certa relação existente entre as duas ações diversas. [...] Prática diametralmente oposta consiste em completar o sentido de certas frases com o infinitivo, antepondo-lhe a partícula de, sem que tenha função apreciável.

Com esta breve revisão percebe-se que, ainda que estas estruturas já tenham sido registradas e comentadas na literatura gramatical tradicional, até o momento ainda não foi fornecida uma explicação clara sobre fenômenos envolvidos em tais construções, quais sejam: a introdução de conectores em contextos onde aparentemente não são “necessários” e a variação no uso dos mesmos, a omissão do sujeito do infinitivo, a promoção do complemento do infinitivo para uma posição tópica, e por fim, o próprio estatuto gramatical deste tipo de oração.

## 2.2 Variação e gramaticalização em uma perspectiva funcional

*Today's morphology is yesterday's syntax (Givón).*

*Today's syntax is yesterday's pragmatic discourse (Heine, Claudi, Hünnemeyer).*

Atualmente, a sociolinguística possui um nicho bem definido enquanto disciplina e área de investigação de fenômenos relacionados à linguagem. Passadas quatro décadas dos trabalhos seminais de William Labov na década de sessenta, hoje é indiscutível que os fenômenos da variação e da mudança são inerentes às línguas, e sabe-se também que estes fenômenos não podem ser inteiramente compreendidos se não forem considerados os

falantes, usuários da língua.

Os estudos da variação lingüística, tais como são realizados hoje, tiveram suas bases fundadas por Weinreich, Labov e Herzog (1978), cuja proposta para um modelo de linguagem procura abranger fatos lingüísticos de uso variável e seus determinantes sociais e estilísticos, considerados através da observação de sua ocorrência na produção lingüística espontânea, e do emprego de um método de análise dos dados assim obtidos. A teoria da variação e mudança lingüística daí resultante rompe com a idéia de estrutura (gramática) homogênea, formada exclusivamente por regras categóricas, que parecem invariavelmente falhas quando confrontadas com a diversidade de formas lingüísticas que denotam um “mesmo significado”, ou “estado de coisas”. Esta ruptura permitiu novas possibilidades de análise e descrição das particularidades no uso destas formas lingüísticas existentes na língua de uma determinada comunidade, como também levou a uma melhor compreensão sobre como ocorre a transformação da própria gramática no curso do tempo.

Se por um lado as bases da sociolingüística estão bem fundamentadas, cabe aos pesquisadores dos dias atuais a tarefa de lançar e testar teorias que possam explicar como a variação ocorre, e de que forma esta dá lugar a uma mudança lingüística. Algumas abordagens que ganharam força mais recentemente no campo das teorias lingüísticas, quais sejam, as teorias funcionalistas, vieram romper ainda mais radicalmente com a noção de gramática/estrutura homogênea, e, por este motivo, formam uma parceria muito promissora no estudo da variação e mudança lingüísticas.

Em uma das correntes mais radicais, a gramática enquanto estrutura consolidada é uma impossibilidade, devido ao constante movimento em que se encontram as formas lingüísticas e dado o surgimento constante de novas formas. A gramática, então, não é um sistema bem-definido, ao contrário, toda a estrutura lingüística é dinâmica e emergente. Relações gramaticais se estabelecem a partir do uso, sem que jamais cheguem a uma completa consolidação (HOPPER, 1987). A variação, então, é uma propriedade inerente às línguas, já que nesse caminho para a estrutura novas formas se sobrepõem às mais antigas em camadas, que gradualmente emergem, num movimento constante (HOPPER, TRAUGOTT 1993).

Adotando uma posição menos radical, Givón (2001a) admite que atuam na língua fatores que são tanto de natureza estrutural (regras consolidadas) quanto motivados pelo

uso, ou seja, mais ligados ao componente pragmático da linguagem. Mas de que maneira o uso modifica formas lingüísticas? Tal pergunta pode ser respondida com relativa facilidade se a mudança em questão se limitar ao plano fonológico, mas se torna mais difícil de ser respondida na medida em que o escopo de análise vai se ampliando para a morfossintaxe e o discurso. Uma tentativa de explicação que (teoricamente) abrange esse tipo de fenômeno é a gramaticalização.

O conceito de gramaticalização é amplamente discutido na literatura lingüística atual, havendo inclusive controvérsias com relação ao seu significado e aplicabilidade. Alguns lingüistas consideram a gramaticalização como apenas um entre vários processos envolvidos na mudança lingüística, ao passo que outros atribuem-lhe o estatuto de uma teoria mais geral da linguagem (NEWMYER 2001; CAMPBELL 2001).

A definição clássica de gramaticalização remonta a Meillet, no início do século XX, que propõe que a gramaticalização é o processo pelo qual itens lexicais autônomos se tornam elementos gramaticais (CAMPBELL, JANDA 2001), como por exemplo, no português, a mudança na expressão do futuro de uma locução verbal (*hei de compreender*) para uma marcação morfológica (*compreenderei*). Aqui cabe questionar se é possível analisar fenômenos de natureza sintática mais ampla a partir desta perspectiva, tais como questões relacionadas à fixação da ordem dos constituintes de uma oração. No caso das sentenças estudadas aqui, é preciso avaliar se há programas de pesquisa em gramaticalização que abonem considerar que as diferentes configurações da sentença representam casos de um processo de gramaticalização, e em caso positivo, qual o caminho esta deve percorrer. Mais especificamente, procura-se compreender se os casos de alçamento (ex.: **O Brasil** é difícil de conhecer) constituem uma forma mais gramaticalizada que os casos de ordenação canônica, em que o alçamento não ocorre (ex.: É difícil conhecer **o Brasil**), tendo em vista que a estrutura das construções de alçamento pode ser analisada como uma oração simples (no exemplo, Brasil funciona como sujeito/tópico da oração e o verbo infinitivo como complemento nominal), ao passo que a sentença canônica permite interpretar o verbo infinitivo seguido de seu complemento como um constituinte oracional.

Cada vez mais pesquisadores da área têm assumido que processos de mudança morfossintática também podem ser explicados através da gramaticalização. Essa ampliação no escopo do conceito de gramaticalização faz com que se passe a aceitar dois tipos de

mudança que podem decorrer da gramaticalização (TRAUGOTT, HEINE 1991, p. 2):

a) Uma que segue a definição clássica de gramaticalização, do tipo:

item lexical > morfema gramatical.

b) Uma que introduz o componente pragmático como origem de estruturas gramaticais do, do tipo:

estrutura do discurso > marcação morfossintática.

Dentro da linha apontada em *b)*, além de itens lexicais, construções sintáticas também podem "assumir funções referentes à organização interna do discurso ou a estratégias comunicativas" por conta da gramaticalização, fazendo com que a construção seja mais fixa e previsível (MARTELOTTA et. al. 1996, p. 45). A gramaticalização pode ser compreendida como um processo de emergência de novas estruturas morfossintáticas através do tempo, podendo ser derivada de precursores sintáticos, lexicais ou pragmático-discursivos, decorrendo do uso real da língua e dependente da frequência de uso para se estabelecer (GIVÓN 1995; HOPPER, TRAUGOTT 1993). A frequência, então, passa a assumir um papel importante no estabelecimento de um processo de gramaticalização.

Para Bybee (2003), uma sequência de palavras utilizada com muita frequência pode tornar-se uma única unidade no processamento, ou seja, quanto mais um determinado conjunto de itens lexicais for usado em uma mesma ordem, mais esta sequência se torna cognitivamente previsível, e mais está sujeita a sofrer outros processos que geralmente acompanham a gramaticalização, como perda de massa fônica e esvaziamento semântico, adquirindo, depois de um longo período de tempo, funções cada vez mais exclusivamente gramaticais. No entanto, parece bastante complicado estabelecer que padrões uma sequência de itens deve apresentar para poder ser considerada dentro desta perspectiva, e também parece ser complexo definir o que caracteriza uma "unidade de processamento", bem como realizar os testes laboratoriais necessários para atestar tal afirmação, problemas que fogem ao escopo e ao tempo de execução deste trabalho. Tendo isso em mente, a posição de Bybee (2003) não é tomada aqui em toda a sua complexidade, mas leva à reflexão de que a ocorrência de certos itens lexicais em colocações frequentes pode ser uma evidência de gramaticalização incipiente (HOPPER, 1991), ou ao menos a condição necessária para que um tal processo tenha início.



Com base nesta abertura teórica, considera-se possível o estudo de fenômenos relacionados à ordem e à combinação de orações como um processo de gramaticalização, o que implica em concordar com a tese de que a sintaxe de hoje é a pragmática discursiva de ontem, ou ao menos, reconhecer que relações morfossintáticas, em muitos casos, se estabelecem como regras a partir de usos pragmaticamente motivados, seguindo o percurso proposto por Givón (1979 apud TRAUGOTT, HEINE 1991) discurso > sintaxe > morfologia > morfofonêmica > zero. Citando Heine e colaboradores, esta linha de pesquisa abriu uma nova janela nos estudos de gramaticalização, que encoraja uma visão de gramaticalização não apenas como a reanálise de material lexical em material gramatical, mas também como a reanálise de padrões do discurso em padrões gramaticais, e de funções do nível do discurso em funções semânticas do nível da sentença<sup>9</sup> (HEINE et al 1991, p. 13).

Este processo passa, primeiramente, pela reanálise de uma estrutura gramatical, e pela fixação da ordem de seqüências de itens lingüísticos, através do aumento da freqüência de uso das mesmas. A reanálise modifica a estrutura gramatical subjacente de uma construção (constituintes, estrutura hierárquica, categorias gramaticais), sem no entanto modificar a manifestação da superfície (marcação morfológica, ordem das palavras, dependendo da possibilidade de mais de uma análise de determinada construção se tornar disponível (CAMPBELL 2001). A reanálise freqüentemente é o primeiro passo de um processo de gramaticalização, ainda que estes dois conceitos não possam ser confundidos, já que é possível haver reanálise, sem que ocorra gramaticalização. No caso das construções de alçamento em estudo, considera-se a possibilidade de reanálise do complemento do verbo infinitivo topicalizado como o sujeito da oração matriz, o que modifica a estrutura hierárquica da sentença, mas não altera a forma como é enunciada.

Além da gramaticalização, pode-se considerar a atuação de outras motivações para o surgimento de novas estruturas em relação à forma de organização do discurso. Uma abordagem funcional da gramática considera a iconicidade um princípio fundamental na estabilização de formas lingüísticas. Uma representação icônica representa uma relação de um-para-um entre formas lingüísticas e funções gramaticais, sem com isso negar a arbitrariedade do sistema de representação lingüístico. Segundo o princípio da iconicidade, a

---

9 Traduzido por mim do inglês: *This line of research has opened a new window on grammaticalization studies, that encourages a view of grammaticalization not simply as the 'reanalysis of lexical as grammatical material' but also as the reanalysis of discourse patterns as grammatical patterns and of discourse-level functions as sentence-level, semantic functions.*

língua se desenvolve de modo a existir um equilíbrio na associação de formas e funções (VOTRE, 1992). Também são consideradas icônicas codificações lingüísticas que mantenham uma vinculação lógica com eventos relatados, como por exemplo, em uma narrativa, os acontecimentos serem narrados na mesma sucessão cronológica em que ocorreram. A iconicidade da gramática, entretanto, não é absoluta, mas apresenta graus, sendo que na maioria das construções mecanismos mais icônicos combinam-se com mecanismos mais convencionalizados, não-transparentes, já que a língua está sempre sujeita a pressões externas e internas que fazem emergir novas relações de formas e funções (GIVÓN, 1995; 2001a). Givón (2001a, p. 35) apresenta alguns princípios<sup>10</sup> de organização gramatical icônica (ou 'regras' da proto-gramática), dentre os quais nos interessam aqui:

(i) Regras de espacialização:

- a. Proximidade e relevância: "Fatias de informação que conceitualmente estão juntas, são mantidas em proximidade espaço-temporal".
- b. Proximidade e escopo: "Operadores funcionais são mantidos mais próximos dos operandos aos quais são relevantes".

(ii) Regras de seqüência:

- a. Ordem e importância: "Uma fatia de informação mais importante é deslocada à esquerda".

Estes princípios norteiam a análise de fatores discursivos na ordenação dos constituintes na construção em estudo. As regras de espacialização e seqüência serviram de base para as hipóteses de natureza discursiva. Estas regras compõem o gatilho pragmático-discursivo que, supõe-se, está por trás das diferentes formas de configuração e de conexão entre as orações.

A partir desta proposta pode-se relacionar os contextos de ocorrência das formas variantes e averiguar em que medida alguns contextos irão favorecer uma forma em detrimento de outra, o que pode resultar na fixação de novas "regras" de ordenação dos constituintes das sentenças em estudo.

---

10 Givón (1995) considera que os princípios não representam leis invioláveis, mas sim predições para a freqüência de distribuição dos elementos na cadeia da fala.

### 2.3 Combinação de orações

Investigando sentenças complexas, percebe-se que há diversas estratégias de combinação de orações e que, dependendo da estratégia adotada, as relações estabelecidas entre as orações se modificam, sendo que nem sempre serão de simples dependência ou independência entre as orações. Isto se deve ao fato de que, entre as sentenças complexas nas quais há uma relação de dependência de uma sobre a outra (seja por compartilhamento de constituintes, seja por que uma é um dos constituintes da outra), existem diferentes tipos de relações possíveis, que se refletem em diferentes restrições quanto à ordenação dessas orações. Ao passo que alguns tipos de oração marginal apresentam grande variabilidade posicional, ou seja, podem ser utilizadas em praticamente qualquer posição em relação à matriz (inicial, medial ou final), outras só podem aparecer em certas posições específicas.

Em face desta complexidade, os conceitos tradicionais de coordenação (orações de um mesmo período que não possuem relações sintáticas uma sobre as outras) e subordinação (orações de um mesmo período que possuem relações sintáticas entre si) são tornam insuficientes para compreender as relações estruturais das sentenças complexas.

Há algum tempo vêm os pesquisadores de constituição funcionalista (inclusive no Brasil) desenvolvendo estudos sobre os processos de constituição do enunciado, os quais mostram à exaustão que o rótulo subordinação, colocado pela tradição (e pela Nomenclatura Gramatical Brasileira) nesse grande bloco de construções complexas, não pode pretender-se configurador de um estatuto único para o bloco (NEVES, 2001, p. 14).

Primeiramente, é preciso diferenciar as sentenças complexas formadas por duas ou mais orações que têm um estatuto independente, daquelas formadas por orações que apresentam interdependência de uma em relação à outra. Por orações de estatuto independente entende-se aqui orações que não compartilham constituintes com outra oração, ou que não possuem uma outra oração como um de seus constituintes.

As orações combinadas com estatuto independente sofrem restrições nos padrões de combinação com outras no que diz respeito apenas a fatores relativos à organização da informação, geralmente relacionados à coesão e coerência textuais. A combinação de orações com estatuto independente é chamada de parataxe, também conhecida como coordenação pela nomenclatura tradicional. A ordenação de orações em uma parataxe não está sujeita a pressões de ordem gramatical, visto que não há compartilhamento de elementos entre as

orações neste caso. É bom ter em mente que, em uma parataxe, a ordem das orações que a constituem está, sim, sujeita a pressões pragmático-discursivas, impulsionadas por princípios icônicos reguladores do fluxo da informação, tais como: estrutura tema-rema (informação já conhecida precede a informação nova), ordenação temporal (dos eventos mais remotos para os mais recentes) e esquemas textuais de figura e fundo.

Sentenças complexas formadas por orações que estabelecem relações mais estreitas, de grande interdependência semântica e estrutural são formadas por uma oração matriz (termo dominante, núcleo) e uma ou mais orações combinadas (termo dependente, margem ou satélite), que podem ser hipotáticas ou encaixadas. Dentre as orações que, combinadas, não apresentam estatuto independente em relação umas às outras, existem aquelas que apresentam restrições posicionais mais frouxas, podendo ocorrer com certa liberdade seja nas marginais da matriz ou intercaladas a esta; e aquelas que apresentam relações mais estreitas de interdependência, com compartilhamento de constituintes e restrições posicionais mais rigorosas, caracterizando-se estas mesmas como um constituinte da oração principal. Ao primeiro tipo de combinação dá-se o nome de hipotaxe (como exemplo pode-se citar as orações adverbiais temporais), e a este último, encaixamento (por ex., as orações substantivas subjetivas e objetivas diretas). Estas duas últimas formas de combinação são conhecidas na nomenclatura gramatical tradicional pelo nome de subordinação.

A noção de integração de orações tem a ver com o tipo de relação estabelecida entre as orações, ou seja, se de total independência, significando orações pouco integradas (e por extensão, pouco gramaticalizadas) ou se de relativa ou grande interdependência significando orações mais integradas (e mais gramaticalizadas). Como pode-se perceber, é uma instância de muita variação, por este motivo, é proposto um continuum, que parte do maior em direção menor grau de autonomia entre as orações, coincidindo com outro, que vai de um grau baixo a um grau muito alto de integração entre as orações (LEHMANN 1988; NEVES 2001; BRAGA 2001).

Outra visão sobre a classificação das sentenças complexas de acordo com diferentes estágios de integração, esboçando um continuum que vai das sentenças combinadas menos integradas para as mais integradas, é caracterizando pelo cline: parataxe > hipotaxe > subordinação. A parataxe se diferencia da subordinação por esta última ser caracterizada pela total inclusão da oração combinada (oração margem) no núcleo (oração matriz), tanto a

hipotática quanto a subordinada são termos dependentes, mas é apenas esta última que é um argumento oracional da oração dominante (HOPPER, TRAUGOTT 1993; BRAGA 2001).

Em qualquer das propostas, no início do continuum não há uma relação de hierarquia entre as orações da sentença complexa, já ao final, há uma relação hierárquica clara. Na proposta de Lehmann (1988) a oração dependente pode sofrer um processo de dessentencialização que pode chegar a tal ponto em que esta se torna um constituinte bem definido dentro da oração principal. A dessentencialização é caracterizada principalmente pela perda de propriedades do verbo da oração dependente, uma vez que atinge o estado de encaixada, com perda da flexão (redução) e pela perda do sujeito, o que caracteriza um grau de integração máximo (LEHMANN, 1988; GORSKI 2001).

Teoricamente, este continuum deve refletir o percurso de mudança lingüística na combinação de orações, mediante um processo de gramaticalização, motivado cognitivamente por princípios de iconicidade diagramática, correspondentes ao paralelismo entre forma e função. As orações que codificam eventos/estados serão mais gramaticalmente integradas quanto mais estes eventos/estados estiverem semântica ou pragmaticamente integrados (GIVÓN 1991, p. 826).

Para distribuir as orações combinadas numa escala de integração, são utilizados parâmetros semântico-sintáticos, os quais controlam, por exemplo, o constituinte da matriz ao qual a oração se vincula e o nível de vinculação sintática estabelecido; a ordem da oração marginal face à matriz; o grau de expansão/redução, indicado pela morfologia verbal e pelo comportamento do sujeito - apagado ou convertido em oblíquo; o grau de entrelaçamento, avaliado pelo compartilhamento de elementos, pela explicitude do conector e pela quantidade de material intermediário (LEHMANN, 1988). Segundo esses parâmetros, uma sentença complexa será tanto mais integrada quanto mais a oração marginal apresentar:

- a) vinculação sintática a constituinte da oração matriz;
- b) variabilidade posicional restrita;
- c) morfologia verbal nominal (forma reduzida);
- d) sujeito não expresso; e
- e) ausência de conector.

O dispositivo de conexão e subordinação tende a ser explícito no início da gramaticalização de uma oração subordinada, sendo gradualmente reduzido a zero. Quando

a cláusula subordinada está ligada à matriz por meio de um conector explícito, isto determina a relação sintática da subordinada em relação à principal (LEHMANN 1988).

Tomando como base esse referencial teórico, busca-se verificar o quão integradas são as orações em estudo, sejam na configuração canônica ou com o complemento do verbo infinitivo alçado. Teoricamente, construções mais integradas se relacionam a estruturas mais gramaticalizadas; assim, pretende-se discutir em que medida diferentes configurações podem constituir ou ser decorrentes de um processo de gramaticalização, tal como definida acima. Por este motivo decidiu-se estabelecer as diferentes configurações da sentença (É difícil conhecer o Brasil / O Brasil é difícil de conhecer / Conhecer o Brasil é difícil), bem como a ocorrência de outros elementos nestas orações, como a presença de sujeito do verbo infinitivo, marcadores discursivos e outros, de modo a esclarecer qual seriam os graus de entrelaçamento possíveis nestes casos.

## **2.4 Topicalidade e alçamento do SN**

A premissa de que a sentença que apresenta o sintagma nominal (complemento do verbo infinitivo) alçada (O Brasil é difícil de conhecer) pode ser considerada variante da sentença canônica (É difícil conhecer o Brasil) está baseada na análise gramatical apresentada por Givón (2001b) para sentenças complexas do inglês, caracterizadas por serem construções de tópico marcado. Devido à semelhança semântica e estrutural das sentenças descritas no trabalho de Givón (2001b) com aquelas analisadas aqui, a proposta deste autor se presta também para a análise dessas sentenças complexas do português.

De acordo com Givón (2001b), a marcação de um sintagma nominal como sujeito (tópico primário) ou objeto (tópico secundário) tem a ver com o uso pragmático da ordenação de palavras no discurso. Assim, uma das motivações para a ordenação dos constituintes em uma sentença será o grau de topicalidade dos referentes da sentença.

Segundo o autor, a informação proposicional codificada em sentenças de estado/evento diz respeito tipicamente a algum "participante-tópico" no estado/evento codificado pela sentença. Tais "participantes-tópicos" são mais comumente entidades codificadas como sintagmas nominais (SNs) do que verbos (eventos) ou adjetivos (estados).

Quando eventos como um todo são topicalizados, esses são "quase sempre nominalizados" (GIVÓN 2001b, p. 253-254). Ainda de acordo com o autor, a topicalidade não é uma propriedade dos referentes no nível da sentença, mas antes uma propriedade dependente do discurso. Isto quer dizer que um participante não é tópico porque foi gramaticalmente codificado como tópico (sujeito, objeto) na sentença isolada, mas sim que os participantes são codificados desta maneira porque são tópicos ao longo de um certo trecho de discurso multi-proposicional. A topicalidade está diretamente relacionada à recorrência do participante em um determinado trecho do discurso. A recorrência se reflete em termos de acessibilidade referencial (recuperação anafórica do referente) e importância temática (persistência catafórica do referente). O tópico, então, é uma entidade não-discreta, à qual pode ser aplicada a noção de escala. O lugar de um item nesta escala de topicalidade irá depender do seu grau de previsibilidade, continuidade e acessibilidade.

A escala de topicalidade, no entanto, não é única, mas deve ser considerada dentro de diferentes dimensões, assim, há a escala na dimensão fonológica (plano de codificação fonológica), a escala na dimensão da ordem das palavras (plano de codificação sintática) e a escala na dimensão dos papéis e animacidade (plano de codificação semântica). Na dimensão fonológica, a escala de topicalidade vai do nível de codificação com maior quantidade de material fônico (sintagmas plenos, pronomes independentes) até chegar à anáfora zero, situação que indica o maior nível de continuidade e acessibilidade do tópico. Na dimensão da ordem das palavras, a escala de topicalidade está expressa na posição relativa do tópico em uma sentença, posições mais à esquerda, que são tradicionalmente as posições da informação já dada, codificam tópicos mais contínuos. Na dimensão dos papéis e animacidade, maior agentividade e animacidade codificam os tópicos mais contínuos (GIVÓN 1983). Observe-se agora a seqüência de sentenças seguinte<sup>11</sup>:

- (6) a. X agrada o João.
- b. Agradar o João é difícil (para X)
- c. É difícil (para X) agradar o João.
- d. O João é difícil (para X) agradar

Na sentença (6d) o sintagma nominal que ocupava a posição de objeto do verbo da oração encaixada (agradar) foi alçado a sujeito do verbo da oração matriz. O alçamento

<sup>11</sup> Adaptada de Givón (2001b, p. 275).

acontece quando o complemento de um verbo em uma sentença complexa passa a se relacionar sintaticamente ao outro verbo, podendo ocorrer o alçamento a sujeito e o alçamento a objeto<sup>12</sup> (GIVÓN 2001b). No caso do alçamento a sujeito, que é o que nos interessa aqui, o elemento alçado mantém uma relação temática com o verbo da oração encaixada.

Givón (2001b, p. 276) considera que alçar um referente de uma oração encaixada para uma posição de argumento gramatical na oração matriz tem o mesmo “efeito pragmático topicalizante” observado em outras regras de promoção, de modo que um referente promovido é mais tópico do que um não promovido. Sustentando esta idéia, o autor chama a atenção para a recorrência da estrutura alçada no gênero notícia de jornal, onde o sintagma nominal alçado é geralmente (GIVÓN 2001b, p. 280):

- a) Tópico global: é mencionado no título como o principal tópico do texto; ou
- b) Tópico local: é mencionado como importante nas orações anteriores à construção alçada.

Com base nisto decidiu-se controlar as características de acessibilidade, permanência e do status informacional do sintagma nominal presente nas orações pesquisadas. Ainda que a construção de alçamento possa diferir da construção com sintagma nominal não alçado em relação ao efeito comunicativo, decidiu-se por tratar ambas as formas como variantes, por se considerar importante averiguar em que medida princípios pragmáticos efetivamente atuam na ordenação de constituintes, e se o contexto discursivo tem algum papel na consolidação da estrutura.

## 2.5 Estudos preliminares

Foram realizadas, como projetos do programa de iniciação científica da Universidade Federal de Santa Catarina, pesquisas que trataram do mesmo tipo de fenômeno que o abordado aqui. Estes primeiros trabalhos tiveram o objetivo mais geral de realizar uma investigação ampla sobre as orações subjetivas reduzidas de infinitivo. Os resultados destes

---

12 É o caso de sentenças como “Ele viu João sair”, em que “João” funciona como argumento dos dois verbos.



estudos foram publicados por Gorski, Mittmann e Neves (2003) e Gorski, Mittmann e May (2004). Serão resumidos aqui os resultados mais relevantes para esta proposta de pesquisa.

As pesquisas realizadas foram de caráter sincrônico e diacrônico. O corpus para a análise sincrônica foi extraído do banco de dados do projeto VARSUL. O primeiro estudo (NEVES, 2003) compreendia apenas as entrevistas da cidade de Florianópolis. No segundo ano da pesquisa (NEVES, 2004), foram incluídas as outras capitais do sul. Para análise diacrônica (MITTMANN, 2003) o primeiro corpus foi formado por 23 textos escritos entre os séculos XVI e XX (com exceção do século XVII, por não haver terem sido encontrados textos adequados deste período), selecionados a partir do critério de apresentarem, na medida do possível, uma linguagem que buscasse incorporar características da língua falada não monitorada, seja pelo tipo de texto (peças teatrais onde estão representados personagens de estratos sociais desprestigiados, tais como empregados domésticos, agricultores, etc.) ou mesmo pelo estilo presente na obra (como p. ex., *O Triste Fim de Policarpo Quaresma*, de Lima Barreto, ou o romance *As vinhas da Ira* do escritor americano John Steinbeck<sup>13</sup>). No segundo ano da pesquisa diacrônica (MAY, 2004) foram acrescentados ao corpus outros textos do português antigo, dos séculos XV e XVII.

A ordenação dos constituintes foi controlada, levando-se em conta os diversos deslocamentos de constituintes para a esquerda do verbo finito. Abaixo, descreve-se rapidamente as diferentes configurações controladas, sob a representação esquemática das configurações da sentença são apresentados exemplos para cada arranjo.

a) Ordem canônica:

SERP3gSUB/ADJg(Prep)(SujInf)INF(Comp/AdInf)

Ex.: É difícil ele comprar uma cerveja.

b) Oração combinada deslocada:

INF(Comp/AdInf)gSERP3gSUB/ADJ

Ex.: Envelhecer é uma merda.

---

13 Na literatura de Lima Barreto encontram-se características próprias do Modernismo, entre elas a busca por representar as personagens brasileiras tais como eram as pessoas brasileiras da época, em seus costumes e em seu falar. Já a obra de Steinbeck retrata a jornada de uma família de camponeses em busca de trabalho, sendo que a tradução utilizada (da década de 40) procurou adaptar a fala dos membros da família para um falar que possuísse características das variedades faladas pelas camadas desprestigiadas da população do Rio Grande do Sul.

c) Complemento/adjunto do infinitivo deslocado:

Comp/AdInfgSERP3gSUB/ADJg(Prep)(SujInf)INF

Ex.: Amigas é muito difícil arranjar hoje em dia, né?

Ex.: Aqui é um lugarzinho bom de se morar.

d) item qualificador deslocado:

SUB/ADJgSERP3 g (SujInf)INF(comp/AdInf)

Ex.: Pior é morar aí em cima.

SERP3: verbo ser conjugado na terceira pessoa do singular

SUB/ADJ: substantivo ou adjetivo

Prep: preposição de ou para.

SujInf: sujeito do verbo infinitivo

INF: verbo infinitivo

Comp/AdInf: complemento do verbo infinitivo ou adjunto adverbial.

Os resultados das frequência para as diferentes configurações da sentença são apresentados na tabela 1, que representa a amostra pesquisada por Mittmann (2003) para os dados de escrita (diacronia) e Neves (2003) para os dados de fala (sincronia).

**Tabela 1**  
**Distribuição do tipo de configuração da sentença nas amostras sincrônica e diacrônica**

Configuração da sentença	Séc. XX fala		Séc. XX escrita		Séc. XIX XVIII		Séc. XVI	
Ordem canônica	36	51%	117	88%	14	70%	06	49%
Oração combinada deslocada	07	10%	04	03%	00	00%	02	17%
Comp. de Infinitivo deslocado	25	36%	04	03%	03	10%	02	17%
Item qualificador deslocado	02	03%	08	06%	04	20%	02	17%
TOTAL	70	100%	133	100%	21	100%	12	100%

Extraída de Gorski, Mittmann e Neves (2003).

A configuração mais freqüente nas duas amostras é a ordem canônica dos constituintes. Na escrita, não foi constatado grande número de deslocamentos, os quais se mostraram mais produtivos na fala. Acredita-se que isto se deve ao fato de o gênero predominante na amostra diacrônica ser o literário (peças teatrais, romances). A pesquisa atual utiliza outro tipo de amostra escrita para que se comparem os resultados.

Em relação ao uso do conector, as tabelas 2 (diacronia) e 3 (sincronia) apresentam as frequências da distribuição deste fator nas amostras de Neves (2004) e May (2004).

**Tabela 2**  
**Distribuição da forma de conexão entre as orações na amostra diacrônica**

Diacronia	Justaposição		de		para	
	apl/tot	%	apl/tot	%	apl/tot	%
Séc. XV	43/91	47%	43/91	47%	05/91	04%
Séc. XVI	24/34	71%	09/34	26%	01/34	03%
Séc. XVII	63/72	88%	05/72	07%	04/72	06%
Sécs. XVIII e XIX	08/12	67%	03/12	24%	01/12	08%
Séc. XX (escrita)	141/152	93%	08/152	05%	03/152	02%
TOTAL	279/361	77%	68/361	19%	14/361	04%

Adaptada de Gorski, Mittmann e May (2003).

**Tabela 3**  
**Distribuição da forma de conexão entre as orações na amostra sincrônica por cidade**

Cidade	Justaposição		de		para	
	apl/tot	%	apl/tot	%	apl/tot	%
Curitiba	12/46	26%	19/46	41%	15/46	33%
Florianópolis	38/64	59%	19/64	30%	07/64	11%
Porto Alegre	05/29	17%	10/29	34%	14/29	48%
TOTAL	55/139	39%	48/139	35%	36/139	26%

Adaptada de Gorski, Mittmann e May (2003).

Pode-se observar uma diferença bastante significativa entre a frequência de uso do conector nas duas amostras. De modo geral, observa-se que o conector foi muito mais utilizado na fala, apresentando uma distribuição bem mais equilibrada do que a que se vê na amostra diacrônica de escrita, mesmo se compararmos apenas o século XX, há uma distribuição muito diferente, com o uso do conector muito restrito no caso da escrita.

Contudo, devemos considerar, em primeiro lugar, que os dados diacrônicos foram extraídos de fontes escritas compondo um corpus bastante heterogêneo, que não compõe uma amostragem ideal, tendo em vista que foi formado por um conjunto de textos aos quais se teve acesso no momento da pesquisa. Em segundo lugar, a escrita costuma ser conservadora quanto ao registro de formas lingüísticas, geralmente sendo caracterizada por alto grau de normatização. Isso pode justificar por que na fala a frequência de uso de combinação com conector seja tão superior à averiguada nas fontes escritas.

Os resultados obtidos com o corpus do banco de dados do projeto VARSUL são interessantes na medida em que apontam para diferenças no uso do conector nas capitais. Os demais fatores sociais (sexo, idade, escolaridade) controlados não se mostraram significativos. Apenas na cidade de Florianópolis foi realizado o controle de três faixas etárias (15-24, 25-49, 50 ou mais), sendo que a construção infinitiva de modo geral se mostrou mais produtiva na fala dos jovens.

O trabalho de Gorski, Mittmann e May (2004) indica que há uma tendência para o uso da preposição *de* em contextos onde ocorre o alçamento do complemento do verbo infinitivo, como no exemplo de Gil Vicente (*Auto da Barca do Inferno*) extraído do corpus diacrônico: “Este foi bom d’embarcar”, ou neste outro, de França Jr (*As doutoras*): “Mas isto é fácil de verificar, senhora [...]”. Esse trabalho também aponta para o uso da justaposição das orações (ausência de conector) quando a estrutura utilizada segue a ordem canônica.

Estes resultados servem de ponto de partida para uma análise mais refinada das sentenças combinadas, norteando a formulação de hipóteses de trabalho e o controle de variáveis lingüísticas e sociais. Os resultados obtidos com o corpus do banco de dados do projeto VARSUL são interessantes na medida em que apontam para diferenças no uso do conector nas capitais. Os demais fatores sociais (sexo, idade, escolaridade) controlados não se mostraram significativos. Apenas na cidade de Florianópolis foi realizado o controle de três faixas etárias (15-24, 25-49, 50 ou mais), sendo que a construção infinitiva de modo geral se mostrou mais produtiva na fala dos jovens.

O trabalho de Gorski, Mittmann e May (2004) indica que há uma tendência para o uso da preposição *de* em contextos onde ocorre o alçamento do objeto do infinitivo para a esquerda do verbo finito. Também aponta para o uso da justaposição das orações (ausência de conector) quando a estrutura utilizada segue a ordem canônica.

Estes resultados servem de ponto de partida para uma análise mais refinada das sentenças combinadas, norteando a formulação de hipóteses de trabalho e o controle de variáveis lingüísticas e sociais.

### 3 PROCEDIMENTOS METODOLÓGICOS

Os métodos utilizados para a investigação se baseiam na sociolingüística variacionista (MOLLICA, 1992) e na lingüística de corpus (SARDINHA, 2004). A primeira contribui com um arsenal metodológico bem estabelecido para o estudo de regras variáveis, e a segunda através de técnicas computacionais de coleta e tabulação de dados de corpora lingüísticos.

#### 3.1 Objetivos

O objetivo geral desta pesquisa é investigar, numa perspectiva variacionista e funcional, o fenômeno da combinação de orações (cf. delimitadas na seção 1.1) em dados de amostras sincrônicas. O enfoque deste estudo recai sobre as diferentes configurações e formas de conexão das construções, utilizando a metodologia da sociolingüística variacionista, bem como métodos não-paramétricos de análise estatística para coleta, organização, investigação e interpretação dos dados.

As questões e hipóteses gerais de trabalho, apresentadas aqui<sup>14</sup>, foram elaboradas levando-se em consideração os resultados de estudos preliminares apresentados na seção 1.2.. Os objetivos específicos são:

- 1) Averiguar os contextos de uso das construções com e sem alçamento do objeto do verbo infinitivo.
- 2) Descrever o uso variável das diferentes formas de conexão entre as orações combinadas detalhadas em 1.1, que são a justaposição, preposição *de* ou preposição *para*.
- 3) Discutir a atuação de princípios funcionais de organização do discurso sobre as

---

<sup>14</sup> Na seção 3.3 “Materiais e métodos” são explicitadas hipóteses mais específicas para os grupos de fatores controlados na pesquisa.

construções de alçamento, e seus reflexos no grau de integração das orações combinadas objeto da pesquisa.

4) Discutir a relação entre frequência de uso, graus de integração e possíveis estágios de gramaticalização das diferentes configurações das construções analisadas neste estudo.

### 3.2 Questões e hipóteses

Investigando as construções de alçamento, pode-se interrogar qual a relação existente entre as diferentes formas de conexão e o emprego de estruturas alçadas, e se relações similares também existem para as estruturas canônicas. Espera-se descobrir em que medida o alçamento se correlaciona a uma determinada forma de conexão, permitindo que se determine mais precisamente o grau de integração da estrutura em estudo. Por outro lado, o alçamento também pode estar relacionado à organização global do discurso, assim, cabe perguntar em que medida fatores como permanência do tópico discursivo, definitude e status informacional do sintagma nominal contribuem para que ocorra ou não o alçamento.

Levantam-se as seguintes hipóteses sobre a ocorrência do alçamento do objeto do verbo finito, e a relação deste fenômeno com a forma de conexão entre as orações empregada, bem como o grau de integração das orações em estudo.

- a) O uso da preposição *de* está relacionado à fixação da estrutura ADJ+DE+INF (o verbo infinitivo não é seguido do seu objeto). Esta estrutura deve se correlacionar a grande frequência de ocorrência das construções com alçamento, indício de fixação e previsibilidade da estrutura ADJ+DE+INF, que pode representar um estágio incipiente de um processo de gramaticalização.
- b) Fatores discursivos concernentes ao sintagma nominal, tais como importância, permanência, acessibilidade e status informacional, entre outros, devem contribuir para o alçamento deste constituinte.
- c) Quando *para* é utilizado, a oração combinada deve manter uma vinculação sintática em relação à matriz como constituinte oracional (irá preservar características de oração substantiva subjetiva ou de oração adverbial final) e, portanto, ter a variabilidade posicional restrita.

- d) Diferentes formas de conexão correspondem a diferentes graus de integração das orações. A estrutura mais integrada é a que apresenta a preposição *de* em construções de alçamento, em segundo lugar as construções conectadas com a preposição *para*, e a estrutura menos integrada é a de justaposição.

Em relação à forma de conexão entre as orações combinadas, é importante refletir sobre a possibilidade de se considerar as três formas (sem conector, *de* ou *para*) como variantes de uma mesma variável, no sentido de Labov (1978), com as três possibilidades veiculando o “mesmo significado”, ou se referindo ao “mesmo estado de coisas”. Em diversas ocorrências, a preposição *para* pode ser interpretada como agregando um traço de finalidade ao enunciado, o que não ocorre com a preposição *de*. Entretanto, em muitos casos preposição *para* não parece exercer esse efeito, podendo ser substituída pela preposição *de* sem perda de conteúdo semântico-pragmático. Vejam-se os exemplos extraídos da fala de um mesmo informante:

(7) Foi duro, né? pra estudar esses quatro anos (VARSUL, CTB 07).

(8) Curitiba ainda é uma cidade boa pra viver (VARSUL, CTB 07).

(9) [...] daí fica duro de terminar a noitada, né? (VARSUL, CTB 07).

(10) Cada vez fica mais fácil de se viver (VARSUL, CTB 07).

Considerando a dificuldade em avaliar até que ponto se deve considerar essas pequenas diferenças no significado semântico do enunciado, decidiu-se optar pelo tratamento das três formas de conexão (justaposição, *de*, *para*) como variantes, já que deste modo será possível isolar os contextos de ocorrência de cada uma das formas, a fim de averiguar se todas essas possibilidades são formas concorrentes para a mesma função. Ainda que o intercâmbio de um conector com outro (*de* e *para*) não pareça alterar o sentido proposicional das sentenças, julga-se mais provável que cada forma de conexão possua uma função própria, já que não há estigma associado ao uso de nenhuma das formas. Por este motivo, *de* e *para* devem apresentar grande predominância de uma forma sobre a outra em contextos específicos:

- e) A preposição *para* deve predominar em contextos nos quais o verbo infinitivo é acompanhado de seu sujeito, nas quais sua função será a de conector entre orações subordinadas.

- f) A preposição *de* deverá predominar nos contextos de alçamento do sintagma nominal. Nestas construções o estatuto de hipotática da oração combinada é duvidoso.
- g) A faixa etária mais jovem de cada recorte da amostra deve utilizar as formas mais marcadas quanto à frequência de ocorrência, que são as sentenças interligadas através de conector (*de* ou *para*). Por conseguinte, a faixa etária mais velha deverá utilizar mais freqüentemente as estruturas mais bem estabelecidas, em que a combinação entre as orações se dá por justaposição (ausência de conector).

### 3.3 Materiais e métodos

São analisadas nesta pesquisa amostras de duas naturezas: língua falada e língua escrita. Para investigar a modalidade falada, é utilizado o banco de dados VARSUL, e para a modalidade escrita, um banco de dados formado a partir do projeto Lacio-Web (ALUISIO et al., 2003), desenvolvido com o suporte do Núcleo Interinstitucional de Lingüística Computacional (PINHEIRO, ALUÍSIO, 2003). Além do conjunto de textos sobre os quais é realizada a análise lingüística, também são utilizados os programas computacionais WordSmith Tools (SCOTT, 1999), Interpretador (ENGESIS; KNIES, COSTA, 1995), GoldVarb 2001 (ROBINSON, LAWRENCE, TAGLIAMONTE, 2001) e SPSS (SPSS, 1999), como ferramentas de trabalho para a coleta e tratamento estatístico dos dados.

O quadro 2 sumariza as características, em termos de número e tipos de palavras, de ambas as amostras, obtidas através do aplicativo *Wordlist* do programa *WordSmith Tools* (SCOTT 1999). O valor expresso em *Tokens* é o número total de palavras que aparecem em todos os textos, *Types* expressa o número de palavras (formas) diferentes que aparecem nos textos. *Type/Token Ratio* é a razão entre o número de palavras diferentes e o número total de palavras dos textos, e *Standardised Type/Token Ratio* é a razão padronizada entre número total de palavras diferentes e o total de palavras, calculada através de razões parciais em trechos de mil palavras. Como é uma medida padronizada, tende a expressar melhor as relações entre formas diferentes e total de palavras na comparação de amostras.



**Quadro 2**  
**Características das amostras de língua falada e escrita**

	Amostra falada	Amostra escrita
Tokens	825.884	51.966.600
Types	21.748	221.011
Type/Token Ratio	2,63 (%)	0,43 (%)
Standardised Type/Token Ratio	32,43 (%)	24,54 (%)

Percebe-se que mesmo a amostra de língua falada sendo consideravelmente menor em números de palavras, esta contém muito mais variedade de formas diferentes do que a amostra escrita. Este resultado pode parecer estranho, visto que o corpus de escrita abrange uma variedade muito maior de textos e assuntos, no entanto, é preciso considerar dois aspectos. Primeiro, a contagem de tipos de palavras do corpus de fala está incluindo anotações e simbologias utilizadas na transcrição da fala que não constituem palavras enunciadas pelas pessoas no momento da entrevista. O segundo aspecto a considerar é que a fala está repleta de interjeições e marcadores discursivos (como “né!”, “tá!”, “ah!”, etc.) que aparecem em abundância em pequenos trechos de fala, mas que não são utilizados nas mesmas condições na escrita. Estes números, portanto, caracterizam as diferenças de ambas as amostras.

### 3.3.1 O banco de dados de fala

O projeto interinstitucional Variação Lingüística Urbana da Região Sul do Brasil (VARSUL) é uma cooperação entre as universidades federais do Paraná, Santa Catarina, Rio Grande do Sul e a PUC-RS, com o objetivo de constituir um banco de dados de língua falada urbana da região sul do Brasil. Atualmente, o banco de dados do projeto conta com 300 entrevistas sociolingüísticas de cerca de uma hora de duração em seu acervo, 24 de cada município incluído no projeto, à exceção de Florianópolis que conta com 36 entrevistas<sup>15</sup>. Os informantes foram estratificados por região, idade, sexo, faixa etária e escolaridade (KNIES, COSTA, 1995). As entrevistas encontram-se gravadas em fita cassete, e as transcrições são disponíveis nas versões impressa e eletrônica, possibilitando que o pesquisador escute, leia

15 As cidades nas quais foram realizadas as entrevistas são: Curitiba, Londrina, Pato Branco e Irati no Paraná; Florianópolis, Lages, Blumenau e Chapecó em Santa Catarina; e Porto Alegre, Flores da Cunha, Panambi e São Borja no Rio Grande do Sul.

ou realize buscas através de aplicativos computacionais.

São utilizadas neste trabalho 24 entrevistas de Curitiba, 24 de Porto Alegre e 36 de Florianópolis. São realizadas duas análises, uma mais geral, com todas as cidades da amostra, considerando as entrevistas que apresentam os mesmos níveis de estratificação social (sexo, idade e escolaridade), e uma somente com a cidade de Florianópolis, na qual são consideradas 12 entrevistas além das já analisadas em conjunto com as demais regiões, que correspondem à faixa etária dos 14 aos 24 anos, já que esta estratificação não está disponível para as outras cidades da amostra.

As entrevistas são identificadas de acordo com uma sigla que corresponde à cidade, seguida de um número cardinal (de 1 a 24), que identifica a ficha social do informante. As abreviaturas utilizadas para as cidades estudadas nesta pesquisa são CTB (Curitiba), FLP (Florianópolis) e POA (Porto Alegre). Apenas em Florianópolis foi adotado neste trabalho um sistema para identificar as entrevistas da faixa etária dos jovens (14 aos 24 anos), que consiste em incluir a letra “J” na identificação da entrevista, antes do número do informante, pois para esta faixa etária foi mantido o mesmo padrão de numeração que vai do número 1 a 24. Assim, uma entrevista identificada como “FLP J3” significa que se trata do informante nº 3 de Florianópolis, da faixa etária dos 14 aos 24 anos.

### **3.3.2 O banco de dados de escrita**

A amostra analisada da modalidade escrita foi montada a partir do banco de dados do Núcleo Interinstitucional de Lingüística Computacional (NILC) (PINHEIRO, ALUÍSIO 2003). O NILC é resultado de uma cooperação interinstitucional entre a Universidade de São Paulo (USP) em São Carlos, a Universidade Federal de São Carlos (UFSCar) e a Universidade Estadual Paulista (UNESP) de Araraquara. O núcleo atua de forma a organizar e disponibilizar corpus da língua brasileira escrita contemporânea para pesquisadores nas áreas da lingüística e computação. O projeto também oferece, através da página do mesmo na internet, ferramentas lingüístico-computacionais, tais como concordanciadores, contadores de frequência e etiquetadores morfossintáticos, através do projeto Lacio-Web. Este projeto conta atualmente com seis diferentes corpus da língua portuguesa, cada qual

organizado de maneira a atender diferentes tipos de pesquisa. Destes, o corpus selecionado para esta pesquisa foi o Lacio-Ref, utilizado como referência para a criação de um sub-corpus, composto apenas dos textos do gênero jornalístico.

O Lacio-Ref é um corpus aberto, composto de textos de diversos gêneros (literário, científico, jurídico,...), tipos (manual, projeto de pesquisa, artigo,...), assuntos (meio-ambiente, esportes, política,...) e formas de distribuição (livros, jornais, internet,...). Os textos têm por característica serem escritos respeitando a norma culta (ALUÍSIO et al, 2003).

Decidiu-se optar pelo gênero jornalístico em função das observações de Givón (2001b) quanto ao uso de construções de alçamento a sujeito neste gênero textual. Os textos que compõem a amostra escrita foram publicados nos veículos Folha de São Paulo e Jornal do Brasil. Os textos deste corpus foram organizados pelo NILC em um sistema próprio de arquivamento e registro, que contém certas peculiaridades. As mais importantes, neste caso, tratam da presença de mais de um texto em um único arquivo, e da falta de padronização dos dados de referência em cada um dos arquivos, assim como da não confiabilidade da data informada em relação à real publicação de cada notícia. A maioria dos arquivos indica a data em que os textos foram coletados, sem que esta seja necessariamente a data de publicação do texto nos veículos, à exceção do jornal *Folha de São Paulo* (PINHEIRO, ALUÍSIO 2003).

Como o corpus Lacio-Ref é aberto, resolveu-se utilizar como identificação das ocorrências o mesmo sistema de nomeação dos arquivos do projeto. Cada arquivo foi identificado de acordo com o seguinte padrão: abreviatura do caderno do qual os artigos fazem parte composta por dois caracteres, seguida da data em que foi coletado o material, a qual é composta por ano (dois últimos dígitos), mês (abreviatura de dois caracteres) e dia (um ou dois dígitos). Os cadernos dos jornais selecionados para esta pesquisa possuem as seguintes abreviaturas:

a) Folha de São Paulo:

- |                    |                   |                    |
|--------------------|-------------------|--------------------|
| ● Agrofolha_____ag | ● Dinheiro_____di | ● Folhinha_____fl  |
| ● Brasil_____br    | ● Empresas_____em | ● Fovest_____fv    |
| ● Ciência_____fc   | ● Especial_____ce | ● Ilustrada_____il |
| ● Cotidiano_____co | ● Esportes_____es |                    |

b) Jornal do Brasil:

- |                    |                      |                     |
|--------------------|----------------------|---------------------|
| ● Caderno B_____cb | ● Imóveis_____im     | ● Folha Teen_____ft |
| ● Cidade_____cd    | ● Informática_____if | ● Tudo_____td       |
| ● Coluna_____cl    | ● Mais_____ma        | ● Turismo_____tu    |
| ● Domingo_____rd   | ● Mundo_____mu       | ● Folha TV_____tv   |
| ● Economia_____ec  | ● Opinião_____op     | ● Veículos_____ve   |
| ● Esportes_____sp  | ● Revista_____re     |                     |

Assim, uma ocorrência identificada como “br94ab28” é proveniente do jornal Folha de São Paulo, do caderno “Brasil”, e a data de publicação/coleta é o dia 28 de abril do ano de 1994.

### 3.3.3 Coleta de dados

Foram coletadas as ocorrências das construções esquematizadas e apresentadas no quadro 1, que apresentem ou não: sintagma nominal alçado, sujeito do verbo infinitivo, outros elementos (advérbios, adjetivos, marcadores discursivos, aposto) em qualquer posição na estrutura. Após a coleta as ocorrências foram codificadas de acordo com variáveis sociais e lingüísticas, de modo a permitir o controle da frequência e dos contextos de ocorrência de cada forma.

Os dados foram coletados com auxílio do pacote *WordSmith Tools* (SCOTT, 1999), em todos os textos que compõem os corpora selecionados para esta pesquisa. A utilização deste programa permite uma maior agilidade e precisão nas coletas de dados. O programa *Concord* do *WordSmith Tools* permite que sejam realizadas buscas em quaisquer bancos de dados devidamente digitalizados em formato de texto (.txt ou .html), através de palavra ou expressão chave. É importante assinalar que a coleta da amostra escrita não foi exaustiva, considerando-se o tamanho da amostra a ser analisada. O programa *Concord* foi configurado para selecionar aleatoriamente 1 a cada 5 ocorrências, até atingir um número máximo de 1.600 ocorrências no arquivo de saída, para cada busca de palavra chave. Deste modo, ainda que a pesquisa na amostra escrita não tenha sido exaustiva, as ocorrências obtidas foram

completamente aleatórias. A amostra de escrita se encontrava previamente etiquetada, o que possibilitou a busca por expressões complexas contendo tanto categorias gramaticais quanto lexemas. Na amostra de fala as buscas foram feitas apenas através de lexemas ou de partes de lexemas. Pelo fato da amostra de fala ser bastante pequena, foi possível realizar buscas exaustivas neste corpus.

Foram coletadas todas as ocorrências obtidas através do concordanciador que se encaixassem na descrição apresentada no quadro 1. Em face do problema de não ser possível realizar uma busca através de uma única palavra chave precisa, já que a variabilidade de formas possíveis para as construções em estudo ser muita, os arquivos de saída geralmente continham muito “lixo”, ou seja, sentenças que não se encaixavam na delimitação do objeto de estudo, ainda que contivessem algum(ns) de seus itens. Por este motivo, cada ocorrência apresentada na saída gerada pelo concordanciador foi analisada, e aquelas que não estivessem de acordo com os critérios estabelecidos na delimitação do objeto foram descartadas. Algumas considerações sobre as ocorrências coletadas:

**a) Expressões de “verbo suporte” na oração encaixada.**

Algumas ocorrências apresentaram, na oração encaixada, expressões que tanto podem ser interpretadas como verbo pleno seguido de seu complemento quanto expressões com os “verbos suporte” *ter*, *dar* e *fazer* (NEVES, 1996), como nos exemplos abaixo.

(11) Daí não foi difícil **dar o corte** no zagueiro para bater de curva [...]. (NILC, sp96ab25).

(12) É difícil **fazer uma previsão**. (NILC, rd96ma06).

(13) Ele avalia que é muito difícil **ter acesso** a todos os delegados do PMDB. (NILC, br94ab28).

Nesses casos decidiu-se por não interpretar tais construções como estruturas de verbo suporte (significando cortar, prever e acessar respectivamente), mas por interpretar os verbos da oração encaixada como plenos. Em primeiro lugar, interpretar tais construções como verbo suporte envolveria o trabalho de pesquisa mais aprofundado sobre esta questão específica, de modo a elaborar critérios mais precisos de inclusão e exclusão deste tipo de ocorrência. Em segundo lugar, estas sentenças parecem aceitar razoavelmente as variações encontradas nas estruturas mais “típicas” em estudo.

(11a) Daí o corte no zagueiro não foi difícil de dar [...].

(12a) Previsão é difícil de fazer.

(13a) Acesso a todos os delegados do PMDB é muito difícil ter.

### b) Expressões "cristalizadas"

Sentenças que apresentassem, na oração encaixada, expressões do tipo dar uma *-ada*, ou seja, o verbo *dar* seguido de artigo indefinido feminino mais um particípio (exs.: *dar uma bolada*, *dar uma escapada*) não foram coletadas, por se considerar que esta estrutura possui um funcionamento particular no discurso, sendo que estas construções já estão integradas a ponto de não apresentarem as variações das sentenças complexas em estudo. Do mesmo modo, expressões idiomáticas tais como *deixar claro* e *ficar de olho* também não foram consideradas ocorrências do fenômeno estudado aqui.

### c) Locuções verbais e verbos auxiliares modais

Foram consideradas neste trabalho todas as ocorrências em que se verificou a existência de locução verbal ou de verbo modal auxiliar. Exemplos:

(14) É melhor **começar a arrumar** as malas (NILC, tu94ag25).

(15) Bah! esse psicotécnico **deve ser** horrível de fazer né? (VARSL, POA 05).

(16) O português **pode ser** uma língua fácil de entender, [...] (NILC, rd96ju23).

## 3.3.4 Variáveis lingüísticas

A variável dependente (VD) e variáveis independentes (VI), também chamadas de grupos de fatores, estão organizados de tal forma que primeiro é fornecida uma descrição da variável/grupo, e a seguir listam-se as variantes/fatores controladas em cada um(a) dos(as) variáveis/grupos seguidas de um exemplo, com o item controlado pela categoria em destaque<sup>16</sup>.

**1) Configuração da sentença:** variável dependente da pesquisa. Neste grupo de fatores são controladas as possíveis formas de ordenação dos constituintes das sentenças coletadas.

16 Os exemplos referidos como "NILC, AC/DC" foram extraídos do corpus do NILC/São Carlos anotado v. 4.5, disponibilizado na internet pelo projeto AC/DC Acesso a Corpora/Disponibilização de Corpora da Linguatca (<http://kkk.linguatca.pt/ACDC/>). Os exemplos extraídos do corpus de referência do NILC estão indicados através do nome do arquivo de onde foram extraídos. Os exemplos extraídos do corpus do VARSUL trazem a indicação da cidade e do número da entrevista na qual foram encontrados.

i. Ordem canônica (estrutura não-alçada)

Ex.: [...] ele leva mais tempo dentro do Correio, preparando, ordenando, do que na entrega. Porque **é difícil aprender a localização da rua** [...] (VARSUL, FLP 14).

ii. Sintagma nominal (SN) objeto do verbo no infinitivo alçado.

Ex.: [...] **o Brasil** não é fácil de conduzir (NILC, AC/DC).

iii. Oração infinitiva deslocada à esquerda.

Ex.: Quem disse que **ganhar dinheiro com literatura no Brasil** é difícil? (NILC, AC/DC).

iv. Apagamento do objeto do infinitivo (o SN foi codificado como anáfora zero)<sup>17</sup>.

Ex.: Como diz seu nome, é fácil de usar **Ø**, mesmo para quem nunca se atreveu a mexer em um computador (NILC, AC/DC).

**2) Forma de conexão:** segunda variável dependente da pesquisa. Objetiva-se investigar a variação entre três possibilidades de conexão entre as orações combinadas em estudo. Como variável dependente, procura-se descobrir se alguma das formas de conexão está mais relacionada a determinada configuração da sentença (a “configuração da sentença” não será utilizada como variável dependente para a forma de conexão).

i. Justaposição (ausência de conector).

Ex.: [...] ainda é difícil **Ø** conhecer o Brasil (NILC, AC/DC).

ii. Preposição *de*.

Ex.: Mas é fácil **de** compreender essas ironias da história (NILC, AC/DC).

iii. Preposição *para*.

Ex.: Depois está difícil também **pra** dar educação, estudo (VARSUL, POA 12).

**3) Forma do objeto do infinitivo:** é esperado que as construções de alçamento, na maioria das vezes, “selecionem” referentes genéricos ou definidos, não devendo ocorrer com referentes indefinidos (GIVÓN, 2001b, p. 276). De acordo com as regras de espacialização e seqüência (GIVÓN, 2001a) também é bastante provável que, quando o complemento do verbo estiver expresso na forma de um pronome anafórico (isso, isto) recuperando um trecho

<sup>17</sup> Nestes casos, as ocorrências serão codificadas com o símbolo de “não se aplica” no momento de realizar as análises estatísticas com o programa GoldVarb 2001.

de discurso, haja maior probabilidade de ocorrência do pronome na posição à esquerda do verbo finito. É variável independente (VI) da “configuração da sentença”.

i. Sintagma nominal genérico.

Ex.: É difícil elas jogar **vôlei** (VARSUL, FLP J21).

ii. Sintagma nominal definido.

Ex.: [...] Depende, **a tramitação no Congresso** é difícil prever [...] (NILC, AC/DC) Ex.: [...] É sempre bom ter **esse tipo de amigo** por perto (NILC, rd96ju23).

iii. Sintagma nominal indefinido.

Ex.: [...] porque é muito bom fazer **um curso universitário**, mas, às vezes, o técnico já, né? [...] (VARSUL, FLP, J01).

iv. Pronomes pessoais, pronomes indefinidos e nomes próprios sem determinante.

Ex.: É difícil de ver **ele** aqui (VARSUL, CTB 05).

Ex.: Trata-se de um projeto com custo elevado. Foi fácil colocá-**lo** em prática em Grenoble? (NILC, cb96fe26).

Ex.: Neste tipo de acidente é difícil culpar **alguém** (NILC, sp96ab09). Ex.: Agora, é fácil expulsar **Edmundo** (NILC, cl96fe13).

v. Pronome resumitivo ou fórico

Ex.: **Isso** aí é bem interessante pra ver (VARSUL, CTB 02).

Ex.: São quinze anos de governos conservadores e **isso** é duro para qualquer um engolir (NILC, ft94ag29).

vi. Objeto do infinitivo apagado (anáfora zero).

Ex.: E tem creme de castanha que também é super fácil de fazer **Ø** (VARSUL, FLP 01).

Ex.: Por ser uma linguagem parecida com C e C++ é muito fácil de se aprender **Ø**. Java é multithread, [...] (NILC, ci96ab16).

**4) Sujeito do verbo no infinitivo:** Considera-se que a expressão do sujeito do verbo infinitivo seja um parâmetro que indica menor grau de integração das sentenças (LEHMANN 1988). Além disso, a ausência do sujeito do verbo infinitivo pode indicar perda de traço verbal, com mais chance para o infinitivo ser analisado como complemento do adjetivo. Também espera-se observar em que medida o sujeito expresso pode estar associado ao uso de conector entre as orações. Por este motivo, esta categoria é variável independente



tanto da “configuração da sentença” quanto da “forma de conexão”.

i. Não expresso.

Ex.: É difícil **Ø** achar pornografia na Internet (NILC, ci96mr05)

ii. Expresso.

Ex.: Agora está difícil de **eu** soltar [gíria] (VARSL, POA 19).

Ex.: Eu acho que daí é mais difícil pra **ti** superar o casamento (VARSL, POA 12). Ex.: Não é difícil **da gente** sentir isso, né? (VARSL, POA 02).

Ex.: [...] é fácil para **uma empresa** aumentar a produção [...] (NILC, AC/DC).

**5) Item verbal:** Foram coletadas todas as ocorrências com qualquer dos verbos estativos ser, estar ou ficar, dos quais o verbo ser deve ser o mais recorrente. É variável dependente da “configuração da sentença”.

i. Verbo ser.

Ex.: Pra começar, **é** triste ver essas crianças na rua pedindo esmola, né? (VARSL, CTB 19).

iii. Verbo estar.

Ex.: Também **está** difícil achar vendedor (NILC, AC/DC).

iv. Verbo ficar.

Ex.: Assim **fica** difícil fazer futebol aqui (NILC, AC/DC).

**6) Item qualificador:** É possível que certos qualificadores ocorram com mais frequência junto de uma forma de conexão em particular, se as distribuições seguirem o mesmo padrão observado nos estudos anteriores, e portanto, esta é variável independente apenas da “forma de conexão”.

i. Formas adjetivas primitivas (em grande maioria os itens fácil, difícil, bom e ruim).

Ex.: [...] não é **fácil** administrar uma cidade (NILC, AC/DC).

Ex.: [...] me ensinaram que era **bom** fazer chá de alface (VARSL, CTB 03).

ii. Formas adjetivas comparativas e superlativas.

Ex: Se não houver segurança é **melhor** adiar a votação (NILC, AC/DC).

Ex: Então leva mais tempo na montagem, à máquina é **pior** pra montar. [tricô] (VARSL, CTB 22).

Ex: Acreditavam os principais integrantes do governo que seria **péssimo** anunciar, de Paris, o nome do novo ministro do Planejamento (NILC, cl96ma30).

iii. Formas substantivas de terminação *-dade*.

Ex.: [...] na época pra ter um skate meu Deus do céu! era **uma dificuldade** (VARUSL, CTB 09).

iv. Formas adjetivas de terminação *-oso* ou *-ado*.

Ex.: É até **perigoso** cair uma árvore, sobre a casa (VARUSL, FLP J18).

Ex.: A TAM não dá desconto porque controlar os descontos é **complicado** (NILC, tu94se15).

v. Formas adjetivas de terminação *-vel*.

Ex.: sem se reforçar, fica **impossível** deter o Flamengo (NILC, sp96ma09).

Ex.: Bah! esse psicotécnico deve ser **horrível** de fazer né? (VARUSL, POA 05).

vi. Formas adjetivas de terminação *-ante* ou *-ente*.

Ex.: É **importante** ter um mapa atualizado (NILC, br94ag15).

Ex.: E também não é **suficiente** dar apenas a terra para um bóia-fria (NILC, rd96ju16).

vii. Formas adjetivas de terminação *-al*.

Ex.: Para quem tem quatro filhos como eu, ter um atlas atualizado é **fundamental** (NILC, br94ag15).

Ex.: [...] é **legal** ter turista no estado (VARUSL, FLP J19)

viii. Expressão nominal complexa.

Ex.: Churrasco não tem grandes mistérios, né? É **uma coisa bem fácil** de se fazer (VARUSL, POA 04).

Ex.: Permitir essa operação que destrói as possibilidades do prazer é totalmente contrário a seus ensinamentos. Mas é **um costume difícil** de combater (NILC, id96ab20).

**7) Presença de material interveniente na sentença:** Assim como a presença de um sujeito do infinitivo, a presença de materiais intervenientes também é parâmetro de menor integração entre as orações (LEHMANN 1988). Neste estudo são considerados materiais intervenientes elementos como advérbios, sintagma nominal, aposto, marcador discursivo, exclamação, expressão locativa ou temporal. Como a integração é uma característica da sentença como um todo, e também está relacionada à forma de entrelaçamento das orações, esta é uma variável independente de ambas as variáveis dependentes.

i. Ausência de material interveniente

Ex.: [...] daí me ensinaram que era bom fazer chá de alface (VARSUL, CTB 03).

Ex.: Além disso, o processo de microimpressão é difícil de reproduzir (NILC, ec96mr26).

ii. Presença de material interveniente

Ex.: O leite era fácil, **né?** de tirar? (VARSUL, CTB 23).

Ex.: E o peixe, **inclusive aqui no Estreito**, é difícil de se encontrar. (VARSUL, FLP 11).

Ex.: Só assim é possível avaliar **por completo** os possíveis danos ambientais (NILC, fc96ma05).

**8) Distância referencial:** distância entre o referente codificado pelo sintagma nominal e sua última menção, contada em número de orações à esquerda, é uma das medidas de continuidade de um referente no discurso propostas por Givón (1983 p. 12-15). A distância referencial é um parâmetro funcional que indica o quanto um referente é acessível ao interlocutor, por isso é definida anaforicamente, a acessibilidade está relacionada ao grau de topicalidade do referente, e portanto deve seguir o princípio funcional da ordem e sequência, também chamado de princípio da ordem linear (ZILLES, 1992; GIVÓN, 1995; GIVÓN, 2001a). Assim, quanto mais acessível, maior a topicalidade. Quanto mais distante estiver a última menção do referente, menos contínuo/acessível será o referente, e menos chances o sintagma nominal tem de ser alçado, ao mesmo tempo, referentes muito contínuos tendem a ser apagados (GIVÓN, 1983). É variável independente apenas da “configuração da sentença”.

i. Primeira menção do participante codificado pelo SN.

Ex.: Charlatanismo ocupa a TV na madrugada

[...]

O capítulo 20 de Antônio Alves, taxista passa às 9 da noite, o capítulo 3 passa às 6 da tarde. Às 8 passa um capítulo inédito de Razão de viver, que é reprisado às 8h30. Em algum horário, no meio dessa confusão toda, passa Colégio Brasil. É impossível acompanhar qualquer **novela do SBT**. (NILC, cl96ma29).

ii. Uma a três orações.<sup>18</sup>

Ex.: Brincava muito de bicicleta, também. Bicicleta eu gostava muito. Mas a bicicleta, naquele tempo,<sup>1</sup> era difícil adquirir **a bicicleta** nova (VARSUL, FLP 04).

iii. Quatro a seis orações.

Ex.: Inalação não combate resfriado, diz pesquisa

<sup>18</sup> Os números subscritos nos exemplos indicam a contagem de orações.

Estudos sugerem que vapor é ineficaz contra vírus

Da Reportagem Local

A inalação de vapor não ajuda<sub>1</sub> a combater o resfriado comum,<sub>2</sub> sugerem dois estudos<sub>3</sub> publicados ontem na revista da Associação Médica Norte-Americana ("Jama").<sub>4</sub>

Recentemente surgiram indícios de que<sub>5</sub> era possível combater o vírus e aliviar os sintomas do resfriado através de inalação de vapor, que aumenta a temperatura interna do nariz. [...] (NILC, br94ab13).

iv. Sete a dez orações.

Ex.: Aulas de esqui são indispensáveis

Da Reportagem Local

Para quem não está habituado<sub>1</sub> a esqui na neve,<sub>2</sub> algumas precauções devem ser tomadas.<sub>3</sub> A primeira delas é<sub>4</sub> levar roupas apropriadas<sub>5</sub> casacos e calças impermeáveis com forro, meias, luvas, gorro e ceroulas de lã são itens básicos.<sub>6</sub>

Para proteger-se do sol,<sub>7</sub> que costuma ser forte nas estações de esqui,<sub>8</sub> leve óculos escuros e filtro solar com fator de proteção acima de 15 para os lábios e a pele.<sub>9</sub> Incluir aulas de esqui no programa é indispensável. Após uma semana de aprendizado, já é possível superar a fase de equilibrar-se nos esquis e ousar algumas manobras. [...] (NILC, tu94ju16).

v. Onze a quinze orações.

Ex.: E: E assim nas festas final de ano, vocês se reúnem?

F: Antigamente reunia,<sub>1</sub> agora está um pouco desligado assim.<sub>2</sub> E: Ah é?

F: Eles vem mais pra cá agora.<sub>3</sub> Mas a gente está ficando preguiçosa<sub>4</sub> eu acho,<sub>5</sub> porque geralmente, porque a família - Ah! Tem um ainda lá em Urubici, né?<sub>6</sub> Geralmente a família era mais unida<sub>7</sub> porque tinha aquele de Urubici, o meu tio de Urubici,<sub>8</sub> que morava ali, né?<sub>9</sub> e aquele de Carianos também que mora<sub>10</sub>, morava ali naquele terreno,<sub>11</sub> então ficava mais fácil de fazer festa com todo mundo, então, reunir assim [...] (VARSUL, FLP J18).

vi. Dezesseis a vinte orações.

Ex.: No final da tarde, estourou em Brasília a notícia, desmentida oficialmente pouco depois, de que a reforma ministerial estava decidida. Luis Carlos Santos iria para a coordenação política, Dorothea Werneck para a Agricultura e Francisco Dornelles para a Indústria e Comércio.

Há quem diga<sub>1</sub> que a fonte do boato<sub>2</sub> foi uma especulação a respeito de possibilidades<sub>3</sub> produzida pelo presidente Fernando Henrique.<sub>4</sub> É difícil acreditar<sub>5</sub> que o Palácio do Planalto<sub>6</sub> não tenha controle sobre uma informação desse tamanho.<sub>7</sub> Mais difícil ainda supor<sub>8</sub> que tenha partido de lá uma deliberada intenção<sub>9</sub> de produzir um balão de ensaio.<sub>10</sub>

O problema é<sub>11</sub> que algo do boato faz sentido.<sub>12</sub> Se finalmente o governo apercebeu-se da necessidade<sub>13</sub> de criar uma instância de articulação política,<sub>14</sub> chegou atrasado a uma conclusão de senso comum.<sub>15</sub> Luis Carlos Santos sempre se bateu pela tese,<sub>16</sub> é um dos líderes cuja avaliação<sub>17</sub> não esconde do governo os perigos iminentes<sub>18</sub> e, por causa disso, desde o final do ano passado cresceu na confiança do presidente.<sub>19</sub>

Faz sentido, portanto.<sub>20</sub>

Dorothea na Agricultura, assim de imediato, fica difícil de entender. Ela própria deu uma gargalhada quando soube da notícia. No PPB produziram-se piadas que recendem a despeito e a desrespeito. [...] (NILC, cl96ab25).

vi. Mais de vinte orações.

Ex.: — O senhor passou anos argumentando que faltava vontade política aos brasileiros para fazer um sério ajuste na economia. Agora que estamos no maior aperto, o senhor diz que deveríamos ter inflação e crescimento. O que mudou?<sub>1</sub>

— Não quis ser polêmico desnecessariamente.<sub>2</sub> Mas o Brasil não está no maior aperto.<sub>3</sub> O país está controlando a inflação por meio de uma moeda muito valorizada. |

<sup>4</sup> Há, portanto, um grande problema não resolvido.<sup>5</sup> O governo substituiu uma coisa por outra.<sup>6</sup> As taxas de juros estão mais baixas do que no ano passado<sup>7</sup> mas continuam altíssimas.<sup>8</sup> Os juros superaltos estão matando a atividade econômica.<sup>9</sup>

— O governo está fazendo a sua parte?<sup>10</sup>

— O ajuste fiscal está sendo muito limitado<sup>11</sup> e este também é um problema<sup>12</sup> que continua sem soluções.<sup>13</sup> Nada foi feito em relação à privatização.<sup>14</sup> Todos falam do leilão da Light,<sup>15</sup> mas onde está o resto?<sup>16</sup> O país não está indo bem no que diz respeito às reformas.<sup>17</sup>

— O Real está em perigo?<sup>18</sup>

— Todo mundo está achando<sup>19</sup> que o trabalho está feito<sup>20</sup> e que tudo vai ficar uma maravilha<sup>21</sup> só porque a inflação foi controlada.<sup>22</sup> Na verdade, esta estabilização foi feita com recursos emprestados.<sup>23</sup> Não adianta controlar a inflação<sup>24</sup> supervalorizando a moeda<sup>25</sup> e jogando as taxas de juros lá em cima.<sup>26</sup> A inflação baixa tem<sup>27</sup> que ser conquistada<sup>28</sup> e deve ser um resultado concreto, e não artificial, das medidas do governo.<sup>29</sup> É por isso que<sup>30</sup> no Brasil é difícil conseguir o crescimento econômico. E é por isso também que o Brasil não consegue ter estabilidade financeira. O mundo continua esperando o Brasil arrumar de vez a sua economia. [...] (NILC, ec96ma23)

**9) Persistência:** funciona como medida de importância do tópico; é usada para controlar quantas vezes o referente aparece no discurso nas dez orações subseqüentes à sua menção (GIVÓN, 1983). Referentes persistentes são entidades importantes no discurso, e por este motivo tendem a ser codificados em posições tópicas. Assim, se um referente for persistente, este tenderia a ser produzido na posição tópica. É controlada contando-se o número de vezes em que o mesmo referente aparece nas próximas dez orações do discurso subseqüente<sup>19</sup>. É variável independente apenas da “configuração da sentença”.

i. Não há retomadas nas próximas dez orações, ou até o final do texto.

Ex.: Cinco dias em Nova York é tempo suficiente para conhecer bons pontos da cidade - especialmente se você vai viajar neste final de ano e está sem companhia.

A cidade, uma das mais agitadas do planeta, sabe como transformar o turista no ser mais sozinho do mundo. Para combater a solidão, a melhor receita é andar a pé -as ruas planas e a geografia é fácil de entender .

Dá para visitar os principais museus, prédios famosos, restaurantes, parques e centros públicos. Metrô, ônibus ou táxi só são indicados quando se quer escapar dos arredores da Quinta Avenida.

Se a visita for bem-organizada, dá até para ir às compras nesse período. Com o real mais forte que o dólar, qualquer conversão para a moeda brasileira resulta em negócio.

LEIA MAIS Sobre Nova York nas págs. 6-6 a 6-8. (NILC, tu94no10).

ii. Uma ou duas retomadas.

Ex.: F: [...] não, hoje a gente vai ao supermercado, né? que a gente vê frutas de vários tipos. Naquele tempo era muito difícil a gente ver um abacate . Pelo menos, aqui no Caieira, era muito difícil a gente ver um abacate<sup>1</sup>. A gente via muito era o caqui, era a goiaba (VARSUL, FLP 12).

iii. Três ou quatro retomadas.

Ex.: F: [...] A gente ia lá pedir e ela não dava daí eu entrava, ia lá por trás lá e roubava, as ameixas.

<sup>19</sup> Esta contagem inclui também as orações em que o referente sofreu apagamento, por considerar que, nestes casos, o referente tem um grau máximo de acessibilidade, sendo facilmente identificado pelo interlocutor.

E: Hum, eu não ligava.

F: Ai! ela não. Maior trabalho, a gente pode pedir, pedir, implorar que era difícil pra ela dar umazinha. Pra ela dar **umazinha**<sub>1</sub> pra gente, a gente tinha que apanhar uma acola de **ameixa**<sub>2</sub> pra ela, pra ela dar **umazinha**<sub>3</sub> pra gente. Era bem assim mesmo (VARSUL, FLP J23).

iv. Cinco ou mais retomadas.

Ex.: F: Está cheio de cocada. **Laranja** é difícil tu encontrares na festa ali.

E: Engraçado, né?

F: É, brinquedo tem, coisa pra vender tem, mas o que encontra é cocada. Não tem nada assim, doce de **laranja**<sub>1</sub>, existe bastante coisa assim de **laranja**<sub>2</sub>, bolo tem de **laranja**<sub>3</sub>, né? sobremesa tem de **laranja**<sub>4</sub>, mas eles não fazem. Só **laranja**<sub>5</sub> assim. E tem muito pouco, [...] (VARSUL, FLP J06).

**10) Status informacional do referente:** Se o referente codificado pelo sintagma nominal for novo, menos chance terá de ser alçado, referentes dados ou inferíveis têm mais chance de serem alçados. Um eferente é considerado novo quando da sua primeira menção no texto, e dado quando já tiver ido mencionado anteriormente. Referentes inferíveis são aqueles que não foram nominalmente mencionados, mas que podem ser inferidos a partir de outros já mencionados, ou seja, possuem uma associação lógica com outros referentes do texto. É variável independente apenas da “configuração da sentença”.

i. Novo.

Ex.: Metheny traz seu garimpo ao Rio, BRAULIO NETO

Guitarrista promete show de 2 horas e diz que **mulheres brasileiras** "não são fáceis de lidar" [...] (NILC, cb96ab22).

ii. Dado.

Ex.: E O que que você acha a respeito, né? dos jovens de hoje, dessa relação de casamento e...

F Ah, estão muito apressadinhos pro meu gosto. Pô, conheceu ontem já quer casar que é Ah, não é assim, né? Acho que a gente tem que conhecer porque apesar que você pode conviver com **uma pessoa** quinze, vinte anos e você não vai conhecer. Você não conhece. Porque a cabeça da pessoa é muito complexa, né? Não sei, é difícil. Cada situação é uma ("questão"), né? Então é difícil você conhecer uma pessoa. Difícil! Você conhece assim superficialmente. Mas conhecer a fundo, você nunca vai conhecer, né? Cada situação é uma [...] (VARSUL, CTB 11).

iii. Inferível.

Ex.: E: E em relação à bebida?

F: Não bebo, não tenho o que falar.

E: Não? E em casa o teu pai também não bebe?

F: Só socialmente ele bebe. É difícil ele comprar uma cerveja, só quando ele vai em festa ou então, ele sai. Depende, em casa mesmo é difícil ele tomar, só quando vai ter festa em casa, aí ele compra. Fora isso é muito difícil (VARSUL, FLP J21).

### 3.3.5 Variáveis sociais

As variáveis sociais serão controladas apenas no corpus do VARSUL, já que este foi montado especialmente para pesquisas sociolinguísticas que lidam com dados de informantes estratificados socialmente.

O corpus de língua falada extraído do banco de dados VARSUL foi dividido em dois diferentes recortes. O primeiro recorte envolve 72 informantes, igualmente distribuídos num quadro de estratificação social de acordo com a cidade de origem, faixa etária, sexo, e nível de escolaridade (quadro 3). Com a estratificação especificada no quadro 3 espera-se reproduzir os resultados de Neves (2004), que apontam para uma diferença no uso do conector em cada uma das cidades do corpus.

**Quadro 3**  
**Distribuição dos informantes em células por cidade, faixa etária, sexo e escolaridade.**

Faixa etária	Sexo	Florianópolis			Curitiba			Porto Alegre			Total
		Pri	Gin	Seg	Pri	Gin	Seg	Pri	Gin	Seg	
25 a 49 anos	Fem	2	2	2	2	2	2	2	2	2	18
	Masc	2	2	2	2	2	2	2	2	2	18
50 anos ou +	Fem	2	2	2	2	2	2	2	2	2	18
	Masc	2	2	2	2	2	2	2	2	2	18
Total de informantes		8	8	8	8	8	8	8	8	8	72

O estudo de Neves (2004) registrou uma recorrência muito maior do uso do conector em Curitiba e Porto Alegre do que em Florianópolis, cidade que apresentou um uso muito mais expressivo da construção sem conector. Se estas diferenças foram mantidas, serão obtidos resultados interessantes no cruzamento destes fatores com as variáveis linguísticas que irão controlar fatores relacionados à organização do discurso.

Com relação às demais variáveis sociais, estas são controladas apenas para confirmar se realmente não há diferenças no uso de diferentes estratégias discursivas por homens e mulheres com maior ou menor escolaridade, já que este foi um aspecto que não foi controlado nos estudos precedentes.

O segundo recorte abrange 36 informantes, apenas da cidade de Florianópolis, também estratificados de acordo com a faixa etária (esta amostra conta com uma faixa etária a mais que a anterior), sexo e nível de escolaridade (quadro 4). Com esta estratificação ilustrada, espera-se refinar o controle dos dados para a variável idade, já que a amostra da

cidade de Florianópolis dispõe de uma faixa etária a mais, sendo esta a principal diferença entre as duas amostras de fala. Por este motivo serão utilizadas 12 entrevistas a mais desta cidade.

**Quadro 4**  
**Distribuição dos informantes de Florianópolis em células por escolaridade e faixa etária e sexo.**

Faixa etária	Sexo	Primário	Ginásio	Secundário	Total de informantes
15 a 24 anos	Fem	2	2	2	6
	Masc	2	2	2	6
25 a 49 anos	Fem	2	2	2	6
	Masc	2	2	2	6
50 anos ou +	Fem	2	2	2	6
	Masc	2	2	2	6
Total de informantes		12	12	12	36

Ainda seguindo a linha dos resultados obtidos por Neves (2004), também se espera que os mais jovens sejam mais produtivos em relação ao uso de algum tipo de conector de modo geral. Em relação ao sexo e escolaridade, o esperado é que estas variáveis não tenham efeito significativo, tanto sobre o alçamento do sintagma nominal quanto sobre a forma de conexão, ainda assim, resolveu-se que seria importante controlar estas variáveis a princípio, realizando os agrupamentos que se julgarem necessário no decorrer da análise.

### 3.3.6 Tratamento estatístico

Todas as ocorrências foram coletadas e codificadas de acordo com os grupos de fatores descritos anteriormente, de modo que seja possível dar entrada dos dados nos programas computacionais utilizados para calcular frequências, pesos relativos e significância para cada variável controlada.

Para a análise multivariada utilizou-se o programa GoldVarb 2001 (ROBINSON, LAWRENCE, TAGLIAMONTE, 2001), que é uma versão mais atualizada do pacote estatístico VARBRUL 2S (PINTZUK, 1988), elaborada para o sistema operacional Windows a partir do programa desenvolvido por Rand e Sankoff (1990) para Macintosh. É o pacote estatístico mais utilizado em pesquisa sociolinguística para a aplicação em regras variáveis. O pacote realiza uma análise multivariada, desde que a variável dependente seja binomial.



Os resultados são apresentados na forma de frequências de distribuição no cruzamento da variável dependente com as demais, e em pesos relativos, que são um índice de aplicação de determinada variante em um dado contexto. O valor do peso relativo pode variar entre 0 e 1, com valores próximos de zero indicando contextos de desfavorecimento da aplicação da regra preestabelecida, e os valores próximos de 1 significando o contexto que mais favorece a aplicação da regra. O Goldvarb utiliza um modelo de interação entre variáveis, selecionando, ao final da análise, aquelas que mais enriquecem o modelo e excluindo as que não, baseando-se em cálculos de *Chi-quadrado* e *Log-Likelihood*.

Também foi realizada análise de dados através do *Statistical Package for the Social Sciences* SPSS (SPSS, 1999), programa desenvolvido especialmente para a estatística na área das ciências humanas. O SPSS possui uma ampla gama de testes estatísticos não paramétricos, dos quais foram utilizados neste trabalho um teste de adequação do modelo e dois testes de associação.

O teste de adequação utilizado foi o *Likelihood Ratio*, que é um teste para dados categoriais que utiliza um modelo *log-linear*. A vantagem deste teste sobre o *Chi-quadrado* de Pearson é que este teste se mostra mais robusto para distribuições com células que contêm valores muito baixos.

Os testes de associação utilizados foram o V de Cramer e a Tau de Goodman e Kruskal. O V de Cramer é um coeficiente de correlação que indica a relação entre duas variáveis, seus valores vão de -1 a 1, sendo que 0 indica que não há nenhuma relação entre as variáveis, e os valores -1 ou 1 indicam que há uma relação perfeita (SEAMAN, 2001). Assim, quanto mais próximos a 1 estiverem os valores, mais poder explicativo a variável independente tem sobre a variável dependente. Esta medida foi selecionada por poder ser aplicada em tabelas maiores do que 2X2, ainda que seus valores possam flutuar de acordo com o número de níveis em cada variável.

A Tau de Goodman e Kruskal também é uma medida de associação aplicável para dados categoriais distribuídos em tabulações cruzadas maiores que 2X2. Esta medida se baseia em redução proporcional de erro, que pode ser interpretada como a variação na proporção de predições incorretas quando as predições para a variável dependente são obtidas considerando: somente a distribuição da variável dependente (probabilidades baseadas apenas nas marginais da linha); a distribuição da variável dependente e das

variáveis independentes (proporções condicionais da linha e da coluna) (BECKER, 1999). Esta característica faz este teste mais confiável nos casos em que a distribuição das ocorrências se mostra alterada. De modo semelhante ao V de Cramer, os valores da Tau de Goodman e Kruskal podem ir de -1 a 1, com o valor 0 representando ausência de associação e os valores (-)1 uma total associação entre as variáveis. Quando a associação é expressa por meio de um número positivo, diz-se que a associação é positiva, o que significa que, quando as ocorrências da variável independente aumentam, também aumentam as ocorrências da variável dependente, pode-se dizer então que o contexto descrito pela variável independente está favorecendo o uso da forma expressa pela variável dependente. Por outro lado, quando a associação é expressa por um número negativo, diz-se que a associação é negativa, o que quer dizer que, quando aumentam as ocorrências da variável independente, diminuem as ocorrências da variável dependente, ou seja, o contexto descrito pela variável independente está inibindo o uso da forma expressa pela variável dependente.

#### 4 RESULTADOS E DISCUSSÃO

Este capítulo foi dividido em quatro seções. Primeiramente são apresentados resultados descritivos de todo o corpus consultado para a principal variável dependente da pesquisa. Chama-se a atenção para algumas peculiaridades desta distribuição, bem como descrevem-se os resultados mais relevantes. Não foram realizados testes estatísticos para as amostras de fala e escrita conjuntamente, por dois motivos: o primeiro diz respeito ao tamanho das amostras, já que a amostra de escrita é muito maior do que a amostra da modalidade falada, considerou-se que quaisquer resultados obtidos somando-se as duas amostras seriam representativos apenas da amostra de fala; o segundo diz respeito aos objetivos da pesquisa, já que o intuito é realizar um diagnóstico da variação da configuração da sentença em cada uma das amostras.

Em seguida apresentam-se os resultados obtidos com a amostra de textos escritos, no que se refere tanto às frequências de distribuição das ocorrências para as diferentes configurações de sentença, quanto para os testes estatísticos aplicados nesta amostra. São exibidos os resultados dos testes de associação no cruzamento de duas variáveis realizados pelo programa SPSS 10.0 e os obtidos pela análise multidimensional do programa GoldVarb 2001.

A terceira parte da apresentação dos resultados se refere aos resultados obtidos com a amostra do VARSUL. Aqui cabe lembrar que o número total de ocorrências desta amostra é de 89 casos, já que a amostra da cidade de Florianópolis utilizada é a mesma, tanto quando foi considerada sozinha quanto quando foi incorporada à amostra geral de fala. A única diferença é que há 28 ocorrências da amostra do VARSUL que se referem à estratificação da faixa etária dos 15 aos 24 anos, que está disponível apenas no banco de dados da cidade de Florianópolis, e que só foram consideradas quando a amostra de Florianópolis foi trabalhada isoladamente.

Assim, na quarta parte, são apresentados os resultados para a amostra de fala apenas

da cidade de Florianópolis. Como o número de ocorrências para esta cidade é de apenas 45 casos, não foi possível realizar análise com o programa GoldVarb 2001, pois o tamanho reduzido da amostra não é suficiente para que se obtenha significância mínima dos testes de *Chi-quadrado* e *Likelihood*. Por este motivo, ainda que tenha sido possível obter alguns resultados significativos nos testes de associação realizados pelo SPSS 10.0, a análise dos dados de Florianópolis será apenas descritiva.

Os resultados são apresentados com a variável dependente “configuração da sentença” ou “forma de conexão entre as orações” ocupando as colunas por questão de economia de espaço. As distribuições foram obtidas através do comando *Crosstabs* do pacote estatístico SPSS 10.0. Apenas para as tabelas que apresentam pesos relativos atribuídos a cada variável foram utilizadas as frequências calculadas pelo programa GoldVarb 2001.

Antes de apresentar a descrição dos dados, é importante tecer algumas considerações sobre as categorias “elipse do SN” e “oração infinitiva deslocada” (na variável dependente “configuração da sentença”) e “elipse” (no grupo de fatores forma do SN), ainda que estas categorias não sejam exaustivamente discutidas com os resultados, já que o foco deste trabalho recai sobre as construções de alçamento e as formas de conexão entre as orações em estudo. As categorias “oração infinitiva deslocada” e “elipse do SN” da variável “configuração da sentença” foram criadas para contabilizar as ocorrências destes tipos (exemplos 17 e 18 respectivamente) objetivando simplesmente uma contagem mais precisa dos dados.

(17) A TAM não dá desconto porque controlar os descontos é complicado (NILC tu94se15).

(18) Na fazenda, uso a tração nas quatro rodas quando chove e **o jipinho** não derrapa. É fácil de manobrar **Ø**, é superlevinho.

Mesmo que estas construções não sejam muito discutidas no que se refere à configuração da sentença, optou-se por incluí-las na análise por conta do fenômeno de variação das formas de conexão.

A tabela 4 mostra as frequências para a variável dependente “configuração da sentença”. Foram coletadas 430 ocorrências ao todo, destas, 41 são construções de alçamento (9,5%), 327 construções canônicas (76%), 27 sentenças de oração combinada anteposta à matriz (6,3%) e 40 ocorrências em que o objeto do verbo infinitivo foi apagado (9,3%).

**Tabela 4**  
**Variável configuração da sentença em toda a amostra (NILC e VARSUL)**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canôn.		Oração inf. des.		Elipse do SN		TOTAL	
		Freq	%	Freq	%	Freq	%	freq	%	freq	%
Forma de conexão entre as orações	justaposição	9	22%	295	90,2%	13	59,1%	8	20%	325	75,6%
	preposição <i>de</i>	30	73,2%	13	4%	0	0%	26	65%	69	16%
	preposição <i>para</i>	2	4,9%	19	5,8%	9	40,9%	6	15%	36	8,4%
Forma do SN (objeto do verbo infinito)	elipse	0	0%	1	0,3%	2	9,1%	39	97,5%	42	9,8%
	SN definido	24	58,5%	154	47,1%	7	31,8%	0	0%	185	43%
	SN genérico	8	19,5%	65	19,9%	4	18,2%	0	0%	77	17,9%
	SN indefinido	0	0%	83	25,4%	8	36,4%	1	2,5%	92	21,4%
	pron. pessoal	3	7,3%	18	5,5%	1	4,5%	0	0%	22	5,1%
	resumitivo/fórico	6	14,6%	6	1,8%	0	0%	0	0%	12	2,8%
Exp. sujeito do infinitivo	não expresso	37	90,2%	289	88,4%	21	95,5%	33	82,5%	380	88,4%
	expresso	4	9,8%	38	11,6%	1	4,5%	7	17,5%	50	11,6%
Presença de materiais interv.	ausência	34	82,9%	297	90,8%	15	68,2%	37	92,5%	383	89,1%
	presença	7	17,1%	30	9,2%	7	31,8%	3	7,5%	47	10,9%
Item verbal	estar	0	0%	16	4,9%	1	4,5%	2	5%	19	4,4%
	ficar	1	2,4%	31	9,5%	2	9,1%	2	5%	36	8,4%
	ser	40	97,6v	280	85,6%	19	86,4%	36	90%	375	87,2%
Forma do elemento nominal	adjs primitivos	38	92,7%	176	53,8%	14	63,6%	28	70%	256	59,5%
	compar e superl	0	0%	20	6,1%	0	0%	2	5%	22	5,1%
	subst. <i>-dade</i>	0	0%	0	0%	1	4,5%	0	0%	1	0,2%
	adjs <i>-oso/-ado</i>	0	0%	9	2,8%	2	9,1%	3	7,5%	14	3,3%
	adjs <i>-vel</i>	1	2,4%	106	32,4%	3	13,6%	3	7,5%	113	26,3%
	adjs <i>-ante/-ente</i>	0	0%	11	3,4%	0	0%	0	0%	11	2,6%
	exp. nom. compl.	2	4,9%	1	0,3%	0	0%	4	10%	7	1,6%
	adjs <i>-al</i>	0	0%	4	1,2%	2	9,1%	0	0%	6	1,4%
Distância da última menção	primeira menção	9	22%	165	50,5%	6	27,3%	1	2,5%	181	42,1%
	1 a 3 orações	19	46,3%	102	31,2%	11	50%	33	82,5%	165	38,4%
	4 a 6 orações	4	9,8%	22	6,7%	2	9,1%	4	10%	32	7,4%
	7 a 10 orações	5	12,2%	12	3,7%	2	0%	0	0%	19	4,4%
	11 a 15 orações	1	2,4%	8	2,4%	0	0%	1	2,5%	10	2,3%
	16 a 20 orações	1	2,4%	4	1,2%	0	0%	0	0%	5	1,2%
	+ de 20 orações	2	4,9%	14	4,3%	1	4,5%	1	2,5%	18	4,2%
Persistência nas 10 orações subsequentes	sem retomadas	18	43,9%	219	67%	14	63,6%	18	45%	269	62,6%
	1 a 2 retomadas	15	36,6%	73	22,3%	8	36,4%	11	27,5%	107	24,9%
	3 a 4 retomadas	5	9,8%	28	8,6%	0	0%	9	22,5%	41	9,5%
	5 a 6 retomadas	2	4,9%	6	1,8%	0	0%	2	5%	10	2,3%
	7a 8 retomadas	2	4,9%	0	0%	0	0%	0	0%	2	0,5%
	9 ou mais retom.	0	0%	1	0,3%	0	0%	0	0%	1	0,2%
Status informacional	dado	31	75,6%	161	49,2%	16	72,7%	38	95%	246	57,2%
	inferível	5	12,2%	71	21,7%	3	13,6%	2	5%	81	18,8%
	novo	5	12,2%	95	29,1%	3	13,6%	0	0%	103	24%
TOTAL		41	100%	327	100%	22	100%	40	100%	430	100%

Uma observação importante sobre a distribuição dos dados é em relação às categorias não-ortogonais “elipse do SN” na variável configuração da sentença e “elipse” na variável forma do sintagma nominal. Apesar de se referirem, a rigor, ao mesmo fenômeno (a omissão do referente), a sobreposição destas duas categorias não é completa. Há duas ocorrências de “elipse” quando a configuração da sentença é “oração infinitiva deslocada à esquerda”. Exemplos:

(19) Dinheiro [pra entrar] pra entrar é difícil, é preciso suar, trabalhar. **Agora pra sair é fácil.** (VARSQL, FLP04).

(20) Inglês [eu sabia] eu sabia ler, mas não entendia o que [que eu estava] que eu estava lendo, entende? Escrever também. Escrever e ler, eu lia, **agora entender foi difícil.** (VARSQL, CTB19).

Em relação à categoria “oração infinitiva deslocada”, a distribuição dos dados se mostrou coerente com as expectativas, na medida em que não foram encontradas ocorrências de oração infinitiva deslocada introduzida pela preposição *de*, ao passo que 40,9% das ocorrências foram introduzidas pela preposição *para* (59,1% de justaposição). Exemplos

(21) Então pra você conseguir digamos não deixar tudo brilhando, mas, pelo menos, **para você conseguir deixar as coisas em ordem é difícil.** (VARSQL, CTB 19).

(22) Mas, às vezes, dispensar uma cerveja fica difícil (NILC, es94de04).

Contudo, como pode ser observado pelas frequências, a preposição *de* tende a ocorrer na maior parte das vezes (65% das ocorrências) em que o sintagma nominal se encontra na forma de elipse. Exemplos:

(23) Imagine, sete cruzeiros, né? [sete] eram sete mil réis, você acredita? **Era difícil de pagar, está?** (VARSQL, CTB14).

(24) Por ser uma linguagem parecida com C e C++ **é muito fácil de se aprender.** Java é multithread, [...]. (NILC, ci96ab16)

Nos cálculos efetuados pelo programa SPSS 10.0 a categoria “elipse do SN” na variável “configuração da sentença” foi mantida, já para a análise realizada pelo programa GoldVarb 2001, as ocorrências deste tipo foram recodificadas, atribuindo-se a estas uma barra (/), que é interpretada pelo programa como um valor que “não se aplica” para a categoria em questão. Deste modo são desconsideradas automaticamente as ocorrências de elipse quando a variável dependente é a configuração da sentença, mas são contabilizadas quando for definida a “forma de conexão” como variável dependente, pois as elipses estão marcadas pela variável “forma do SN”.

Ainda na categoria “elipse do SN” há uma ocorrência codificada na “forma do SN”

como sintagma nominal indefinido. Esta codificação foi feita considerando que os valores “elipse do SN” seriam marcados como “não se aplica” para a análise multidimensional do GoldVarb, sendo atribuída à seguinte ocorrência:

- (25) Se o Sarney, se esses outros aí, Geisel, Figueiredo, que já eram pessoas de idade não souberam fazer nada [pro] pro bem do país, **um cara aí, de trinta e oito, quarenta anos** fica meio difícil pra governar [um] um país, né? (CTB03)

Neste exemplo temos, à primeira vista, um caso de alçamento. Entretanto, o constituinte que está ocupando a posição do sujeito da oração finita não é o objeto do verbo no infinitivo (um país), mas sim o sujeito deste verbo. Reescrevendo a sentença (25) na ordem canônica tem-se:

- (25a) fica meio difícil pra **um cara aí, de trinta e oito, quarenta anos** governar um país, né?

Neste caso, temos o sintagma nominal objeto do verbo no infinitivo codificado como um sintagma nominal indefinido, entretanto a configuração da sentença não se encaixa em nenhuma das categorias previamente definidas. Assim, ao invés de criar uma categoria apenas para esta ocorrência, ou excluí-la da análise, decidiu-se marcar a configuração da sentença como “não se aplica”.

De modo geral, a distribuição das ocorrências nas variáveis consideradas se mostrou de acordo com o esperado, em consonância com os estudos preliminares. A configuração da sentença mais recorrente é a canônica, e a forma de conexão entre as orações mais freqüente se dá pela simples justaposição das orações, e entre os conectores, a preposição *de* é a mais utilizada.

O tipo de item qualificador mais predominante é o adjetivo, especialmente nas formas primitivas (difícil, fácil, bom), ainda que haja um considerável número de adjetivos derivados com o sufixo *-vel* (possível, impossível, agradável, entre outros) ocorrências estas que se mostraram predominantes na amostra de textos escritos (jornal). Exemplos:

- (26) Na aula de química é **difícil** alguém falar alguma coisa (VARSUL, FLP J21).

- (27) A família é **bom** estar reunida , né? (VARSUL, CTB 03).

- (28) Ainda hoje, é **possível** ver a arte da criançada (NILC, cl96fe13).

A tabela 5 apresenta a distribuição das ocorrências para cada uma das amostras consideradas, de textos escritos (jornal) e de fala (VARSUL).

**Tabela 5**  
**Variável configuração da sentença de acordo com a origem das ocorrências (NILC ou VARSUL)**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canônica		Oração infinitiva deslocada		Elipse do SN		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Origem da ocorrência	jornal (NILC)	26	63,4%	269	82,3%	7	31,8%	11	27,5%	313	72,8%
	fala (VARSUL)	15	36,6%	58	17,7%	15	68,2%	29	72,5%	117	27,2%
Total		41	100%	327	100%	22	100%	40	100%	430	100%

Estes resultados refletem a distribuição que era esperada no que se refere ao número de elipses, muito mais freqüentes na fala do que na escrita. Das 41 construções de alçamento a sujeito registradas em toda a amostra, 63,4% encontram-se em textos de jornal, e 36,61% foram encontradas na fala urbana das capitais da Região Sul. Já em relação ao número de alçamentos, esperava-se uma maior recorrência das construções alçadas nos textos de jornal do que na fala, baseando-se nas observações de Givón (2001b), segundo as quais as construções alçadas são mais comuns em textos jornalísticos, não sendo muito comuns no inglês falado.

As construções de alçamento em inglês, entretanto, apresentam variações específicas daquela língua, não sendo possível estabelecer equivalências absolutas entre o inglês e o português. Por exemplo, formas como (28), podem ser comuns no inglês escrito, mas é pouco provável que tenham um equivalente sintático no português (29), ainda que certamente tenham equivalentes semânticos.

(29) *He is thought to be from Moyie Springs.*<sup>20</sup>

(30) \*Ele é pensado ser de Moyie Springs.

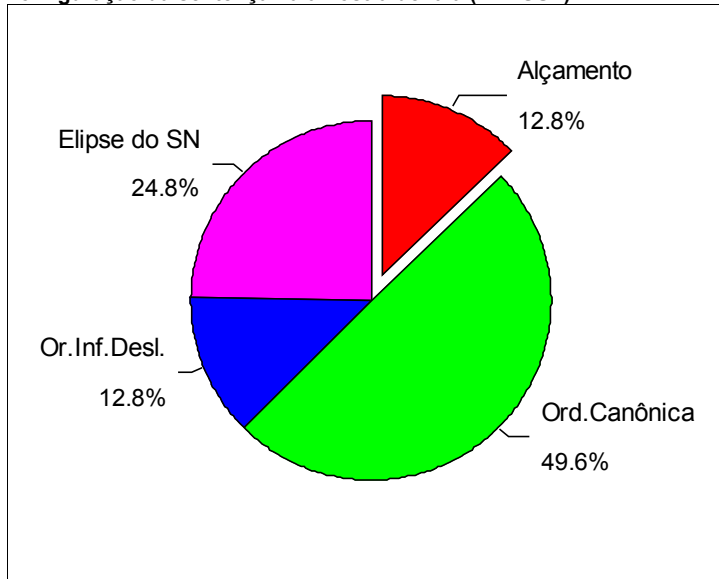
Do mesmo modo, não parece ser possível afirmar que o alçamento a sujeito não é um fenômeno comum no falar do sul do Brasil, pois, ainda que em menor proporção em relação à freqüência de orações na ordem canônica (44,9%), o alçamento ocorreu na amostra do VARSUL em 12,4% dos casos coletados (gráfico 1), o que representa cerca de 1/8 do percentual total de ocorrências da amostra de fala<sup>21</sup>.

20 Extraído de Givón (2001b, p. 280).

21 Está-se considerando aqui o universo de 117 ocorrências coletadas na amostra do VARSUL. Se forem consideradas apenas as 89 ocorrências tomadas para análise da amostra geral de fala, estas proporções ainda se mantêm, com 12,4% de alçamentos, 29,2% de elipses do SN, 13,5% de orações infinitivas deslocadas e 44,9% de orações na ordem canônica.

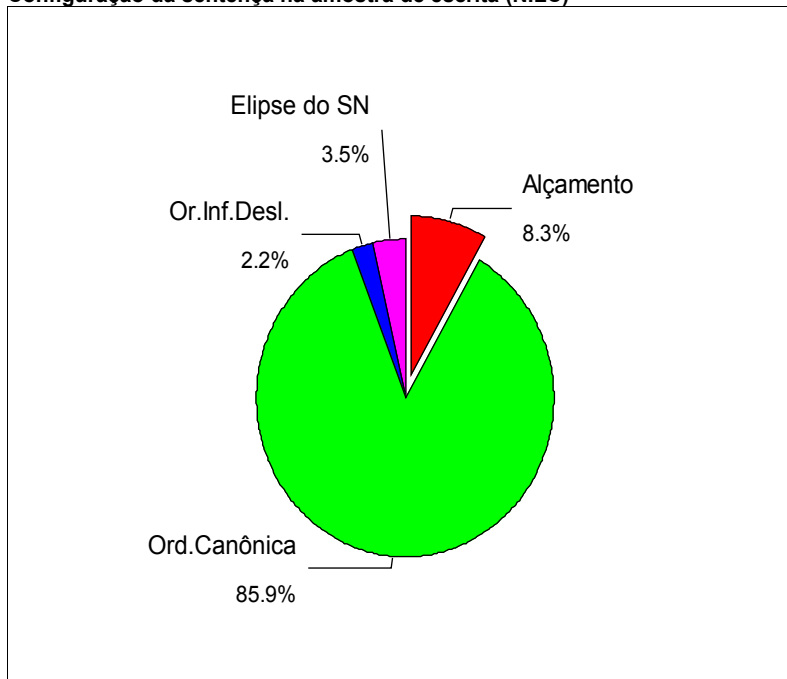


**Gráfico 1**  
**Configuração da sentença na amostra de fala (VARSUL).**



Na amostra de textos de jornal, a ordem canônica representa 85,9% da amostra, ao passo que o alçamento conta com apenas 8,3% das ocorrências, cerca de 1/12 do percentual total (gráfico 2).

**Gráfico 2**  
**Configuração da sentença na amostra de escrita (NILC)**



Assim, se por um lado em relação aos valores absolutos no montante total dos dados o alçamento foi mais recorrente na amostra de jornal do que na amostra de fala, se tomarmos a proporção que as construções de alçamento ocupam em cada amostra separadamente, percebe-se que a oralidade, também no que se refere a estratégias de ordenação dos

constituintes, apresenta variações maiores que a escrita.

#### 4.1 Resultados da amostra de textos escritos (NILC)

Os resultados aqui apresentados estão reunidos em dois grupos, de acordo com a variável considerada como variável dependente. Primeiramente, são apresentadas as frequências absolutas e porcentagens para a distribuição das ocorrências com a “configuração da sentença” como variável dependente, seguidas dos resultados dos testes estatístico aplicados a esta variável. O segundo grupamento de resultados trata da “forma de conexão” tomada como variável dependente. Para a variável dependente “configuração da sentença”, os procedimentos estatísticos foram realizados considerando-se as seguintes variáveis independentes (grupos de fatores):

- a) Forma de conexão entre as orações;
- b) Forma do SN (objeto do verbo infinitivo);
- c) Expressão de sujeito do infinitivo;
- d) Presença de materiais intervenientes;
- f) Item verbal;
- g) Distância para a última menção do referente (SN);
- h) Persistência do referente (SN) nas próximas 10 orações;
- i) Status informacional do referente (SN) na presente oração.

Ao tomar como variável dependente a forma de conexão, foram consideradas as seguintes variáveis independentes:

- a) Forma do SN (objeto do verbo infinitivo);
- b) Expressão de sujeito do infinitivo;
- c) Forma do item qualificador.

São apresentados resultados apenas das variáveis que atingiram a significância mínima ( $\alpha=0,05$ )<sup>22</sup>, seja nos testes do SPSS 10.0 quanto do GoldVarb 2001.

---

22 O programa SPSS 10.0 dispõe de dois testes de significância para variáveis categoriais: o qui-quadrado de Pearson e o *likelihood ratio*. Tomou-se como medida de significância para este estudo apenas os valores calculados pelo *likelihood ratio*, uma vez que as distribuições apresentam um grande número de células com valores esperados menores do que 5.

#### 4.1.1 Configuração da sentença em textos escritos

De modo a reduzir, ou em alguns casos eliminar, células com valores iguais a zero, foram realizados agrupamentos e exclusões de categorias dentro das variáveis:

1) **Variável independente “distância referencial”**: Apenas 28 ocorrências foram encontradas quando a distância referencial era maior do que 10 orações, assim, agruparam-se todas estas ocorrências na categoria “7 a 10 orações”, que passou a representar as ocorrências com distância referencial da última menção equivalente a 7 orações ou mais.

2) **Variável independente “forma do SN”**: Apenas para a análise estatística realizada pelo programa GoldVarb 2001 foram excluídas as ocorrências de sintagma nominal indefinido, pois sua distribuição era igual a zero em construções de alçamento.

A tabela 6 apresenta as frequências para a variável dependente “configuração da sentença” na amostra escrita, que compreende um total de 313 ocorrências, das quais 26 (8,3%) são de construções de alçamento, 269 (85,9%) são sentenças na ordem canônica, 7 (2,2%) são orações combinadas antepostas à oração matriz, e 11 (3,5%) são sentenças em que o objeto do infinitivo foi apagado (elipse do SN).

**Tabela 6**  
**Variável configuração da sentença na amostra de escrita (NILC)**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canônica		Oração infinitiva deslocada		Elipse do SN		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma de conexão entre as orações	ausência	2	7,7%	253	94,1%	6	85,7%	2	18,2%	263	84%
	preposição <i>de</i>	23	88,5%	4	1,5%	0	0%	9	81,8%	36	11,5%
	preposição <i>para</i>	1	3,8%	12	4,5%	1	14,3%	0	0%	14	4,5%
Forma do SN (objeto do verbo infinito)	elipse	0	0%	0	0%	0	0%	11	100%	11	3,5%
	SN definido	18	69,2%	138	51,3%	3	42,9%	0	0%	159	50,8%
	SN genérico	1	3,8%	50	18,6%	1	13,3%	0	0%	52	16,6%
	SN indefinido	0	0%	63	23,4%	3	42,9%	0	0%	66	21,1%
	pronome pessoal	2	7,7%	15	5,6%	0	0%	0	0%	17	5,4%
	resumitivo/fórico	5	19,2%	3	1,1%	0	0%	0	0%	8	2,6%
Exp. do sujeito do infinitivo	não expresso	25	96,2%	255	94,8%	7	100%	11	100%	298	95,2%
	expresso	1	3,8%	14	5,2%	0	0%	0	0%	15	4,8%
Distância da última menção	primeira menção	6	23,1%	145	53,9%	4	57,1%	1	9,1%	156	49,8%
	1 a 3 orações	10	38,5%	74	27,5%	1	14,3%	8	72,7%	96	29,7%
	4 a 6 orações	3	11,5%	17	6,3%	1	14,3%	1	9,1%	22	7%
	+ de 7 orações	7	26,9%	33	12,3%	1	14,3%	1	9,1%	42	13,4%
Status informacional	dado	19	73,1%	123	45,7%	3	42,9%	9	81,8%	154	49,2%
	inferível	3	11,5%	68	25,3%	2	28,6%	2	18,2%	75	24%
	novo	4	15,4%	78	29%	2	28,6%	0	0%	84	26,8%
Total		26	100%	26	100%	7	100%	11	100%	313	100%

#### 4.1.1.1 Alçamento e forma de conexão entre as orações na escrita

A maioria dos alçamentos ocorre concomitantemente à preposição *de* (88,5% dos casos de alçamento, ex. 30), o mesmo ocorrendo quando há a elipse do objeto do verbo infinitivo: 9 em cada 11 elipses ocorrem com a preposição *de* (ex. 31). A preposição *para* não foi muito recorrente na amostra de textos de jornal, contando com apenas 14 ocorrências no total. Exemplos:

(31) Além disso, **o processo de microimpressão é difícil de reproduzir** (NILC, ec96mr26).

(32) Não é bom ter sempre que correr atrás do resultado. Contra o Itaperuna (1 a 0 Vasco, vitória conseguida de pênalti, no último minuto), e se eles fizessem um gol? la ser ruim de virar Ø [o jogo]. Contra o América será o mesmo tipo de jogo (NILC, sp96ab25).

O quadro 5 apresenta os resultados dos testes de significância e associação computados pelo programa SPSS 10.0 para o cruzamento destas variáveis.

**Quadro 5**

**Testes de significância e associação – Configuração da sentença vs. Forma de conexão (escrita)**

<i>Likelihood Ratio</i>	156,343	sig 0,000
Tau de Goodman e Kruskal	0,47	sig 0,000
V de Cramer	0,61	sig 0,000

Estas medidas indicam uma associação positiva entre as variáveis “configuração da sentença” e “forma de conexão”. Estas variáveis apresentaram a associação mais forte obtida nesta amostra para a variável dependente “configuração da sentença”. A diferença observada nas duas medidas de associação é devida aos diferentes procedimentos de cálculo (a Tau de Goodman e Kruskal se baseia em PRE, ao passo que o V de Cramer não).

Os pesos relativos foram calculados pelo programa GoldVarb 2001 considerando-se o alçamento como aplicação da regra em contraste com as demais configurações da sentença (ordem canônica e oração infinitiva deslocada à esquerda somadas), de modo a transformar a variável dependente em uma variável binomial.

De todos os grupos incluídos no modelo, apenas a “forma de conexão entre as orações” (tabela 7) e a “forma do SN” foram selecionados na análise multidimensional do GoldVarb 2001, o que significa que as demais categorias não exercem influência sobre o alçamento do objeto.

**Tabela 7**  
**Alçamento vs. Forma de conexão – pesos relativos da amostra de escrita**

		ALÇAMENTO		Total da amostra		
		Freq.	%	P.R.	freq	%
Forma de conexão entre as orações	ausência	2	7,7%	0.33	261	86,5%
	preposição <i>para</i>	1	3,8%	0.75	14	4,6%
	preposição <i>de</i>	23	88,5%	0.99	27	8,9%
Total		26	100%		302	100%

Input 0,014 Log likelihood -18.465 Sig. 0.001

Estes resultados complementam aqueles obtidos através dos testes de associação. Os valores altos dos pesos relativos indicam alta determinação do alçamento pela forma de conexão utilizada. Neste caso, a preposição *de* favorece fortemente o alçamento a sujeito.

A primeira hipótese de trabalho referente à forma de conexão estabelecia que a recorrência da preposição *de* deveria estar relacionada à estrutura ADJ+DE+INF, em que o verbo infinitivo não é seguido do seu objeto. Esta estrutura estaria correlacionada à grande frequência de ocorrência das construções com alçamento, o que pode indicar o início da fixação da estrutura ADJ+DE+INF.

Observou-se que, apesar da preposição *de* favorecer fortemente o alçamento, como também ser predominante nos contextos de elipse do sintagma nominal, ambos contextos em que o verbo infinitivo se encontra sem seu complemento, a recorrência deste tipo de estrutura foi muito baixa para que se possa considerar que a mesma esteja efetivamente se fixando. No presente momento, a ordem canônica com a justaposição das orações é a principal forma de codificação desta construção.

#### 4.1.1.2 Alçamento e forma do SN na escrita

Em relação à forma do SN, esperava-se que as construções de alçamento ocorressem preferencialmente com referentes genéricos ou definidos, não devendo ocorrer com referentes indefinidos (GIVÓN 2001b p. 276). Também se considerou que, quando o complemento do verbo estivesse na forma de um pronome anafórico (isso, isto), haveria maior probabilidade de ocorrência do pronome na posição alçada. Exemplos:

(33) Depende, **a tramitação no Congresso** é difícil prever (NILC, ci96mr26).

(34) [...] foi uma emoção e **isso** é difícil de explicar (NILC, id96mr16).

Grande parte das ocorrências era de sintagmas nominais definidos, que se distribuíram predominantemente na configuração “ordem canônica”. Ainda assim, a maioria dos objetos alçados a sujeito foram sintagmas nominais definidos (69,2% das ocorrências de alçamento). Os sintagmas indefinidos não ocorreram nunca em posição alçada. O quadro 6 mostra a força da associação entre a configuração da sentença e a forma do SN.

**Quadro 6****Testes de significância e associação – Configuração da sentença vs. Forma do SN (escrita)**

<i>Likelihood Ratio</i>	129,372	sig 0,000
Tau de Goodman e Kruskal	0,32	sig 0,000
V de Cramer	0,62	sig 0,000

Considerando a medida mais conservadora, a força da associação entre estas duas variáveis é moderada. Na tabela 8 vêem-se os valores de pesos relativos para o segundo grupo de fatores selecionado pelo GoldVarb, a “forma do SN”.

**Tabela 8****Alçamento vs. Forma do SN – pesos relativos da amostra de escrita**

		ALÇAMENTO		Total da amostra		
		Freq.	%	P.R.	Freq.	%
Forma do SN	SN genérico	1	3,8%	0.09	52	22%
	pronome pessoal	2	7,7%	0.55	17	7,2%
	SN definido	18	69,2%	0.63	159	67,4%
	resumitivo/fórico	5	19,3%	0.96	8	3,4%
Total		26	100%		236	100%

**Input 0,014 Log likelihood -18.465 Sig. 0.001**

A hipótese inicial se provou parcialmente correta. Sintagmas definidos e pronomes fóricos são fortes candidatos ao alçamento. Contrariamente à previsão de Givón (2001b), sintagmas nominais genéricos parecem desfavorecer em muito o uso da construção alçada. Ainda assim, o número total muito baixo de pronomes resumitivos ou fóricos (8 ocorrências em toda a amostra) faz com o resultado para esta categoria seja interpretado com cautela.

Ainda que estes dois grupos de fatores apresentados (forma de conexão e forma do SN) tenham sido selecionados na análise multidimensional, não foi atingida a convergência nesta rodada estatística. Uma explicação para isso pode ser devido aos valores muito baixos de aplicação da regra (alçamento), em contraste com os de não aplicação (ordem canônica e oração infinitiva deslocada), já que grande parte das ocorrências se encontra na variante ordem canônica. Foi então realizada nova análise multidimensional, desta vez considerando a “ordem canônica” como valor de aplicação da regra.

#### 4.1.1.3 Ordem canônica: forma de conexão e expressão do sujeito na escrita

Na distribuição das frequências apresentada na tabela 3 a ordenação canônica das sentenças se mostrou o ambiente preferencial para a justaposição entre as orações (34).

(35) É por isso que no Brasil é difícil conseguir ~~o~~ o crescimento econômico (NILC, ec96ma23).

Os grupos selecionados durante a análise multidimensional foram a “forma de conexão” (tabela 9) e “expressão de sujeito do verbo infinitivo”.

**Tabela 9**  
**Ordem canônica vs. Forma de conexão – pesos relativos da amostra de escrita**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma de conexão entre as orações	preposição <i>de</i>	4	1,5	0.01	27	8,9
	preposição <i>para</i>	12	4,5	0.04	14	4,6
	ausência	253	94	0.67	261	86,5
Total		269	100%		302	100%

Input 0,950 Log likelihood -50,186 Sig. 0,023

Desta vez a convergência foi atingida, e os resultados obtidos indicam que a justaposição é o contexto lingüístico de combinação de orações que mais favorece a escolha da variante canônica da sentença, ao passo que a presença de um conector a inibe, complementando, então, os resultados apresentados anteriormente para a forma de conexão.

Quando a “expressão do sujeito do infinitivo” foi cruzada à variável dependente “configuração da sentença” no programa SPSS, este não apresentou significância mínima no teste de *Likelihood*, apesar de ter sido selecionado na análise multidimensional do GoldVarb 2001. A presença de sujeito do verbo infinitivo também favorece o uso da ordem canônica (tabela 10). Exemplo:

(36) Não é fácil para ela descrever sua sensação durante a entrega do prêmio (NILC, cb96mr01).

**Tabela 10**  
**Ordem canônica vs. Expressão do sujeito do inf. – pesos relativos da amostra de escrita**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Expressão do sujeito do infinitivo	não expreso	255	94,8%	0.46	287	95%
	expreso	14	5,2%	0.95	15	5%
Total		269	100%		302	100%

Input 0,950 Log likelihood -50,186 Sig. 0,023

#### 4.1.1.4 Distância referencial e status informacional na escrita

Trabalha-se com a idéia de que as construções de alçamento representam uma variante regulada por esquemas globais de organização do discurso, relativos a princípios de iconicidade diagramática. Assim, era esperado que o comportamento do referente no discurso pudesse ser avaliado em variáveis como a distância referencial, a permanência, e o status informacional do referente, e que estas se mostrassem associadas à configuração da sentença.

Na amostra de textos escritos, nenhuma das variáveis discursivas propostas foi selecionada pelo programa GoldVarb 2001 para o alçamento a sujeito. Duas variáveis atingiram significância para esse cruzamento no SPSS: a “distância referencial” (quadro 7) e o “status informacional” (quadro 8), mas ambas apresentaram associações bem fracas com a “configuração da sentença”.

**Quadro 7**

**Testes de significância e associação – Config. da sentença vs. Distância referencial (escrita)**

<i>Likelihood Ratio</i>	22,264	sig 0,008
Tau de Goodman e Kruskal	0,03	sig 0,008
V de Cramer	0,15	sig 0,000

Observando as freqüências da tabela 6, percebe-se que o alçamento ocorre principalmente quando o referente é altamente acessível, 38,5% dos casos de alçamento ocorrem quando o referente alçado foi mencionado no máximo na terceira oração precedente. Estas freqüências vão ao encontro da hipótese de que referentes alçados são mais acessíveis do que os não alçados. Exemplo.

- (37) O KY, gel da Johnson e Johnson, tem base aquosa, é seguro e próprio para ser usado com preservativo. Lançado agora na "embalagem supermercado", vem com três bisnagas de 10 gramas cada. **As bisnaguinhas** são práticas de carregar, na bolsa ou no bolso (NILC, re94no20).

O resultado para a variável “status informacional” é complementar ao da “distância referencial”, já que, para um referente ser acessível, este deve ser dado. A hipótese de pesquisa para esta variável independente era de que referentes dados ou inferíveis teriam mais chance de serem alçados.

**Quadro 8**

**Testes de significância e associação – Config. da sentença vs. Status informacional (escrita)**

<i>Likelihood Ratio</i>	15,696	sig 0,015
Tau de Goodman e Kruskal	0,02	sig 0,001
V de Cramer	0,14	sig 0,045



Esta hipótese se mostra coerente para os referentes “dados”, mas não para os “inferíveis”. Apenas 11,5% das construções de alçamento ocorreram com referentes “inferíveis”, ao passo que 73,1% dos casos de alçamento se deram com um referente “dado”.

#### 4.1.2 Forma de conexão em textos escritos (NILC)

Foram cruzadas as distribuições da “forma de conexão entre as orações” com as variáveis independentes: “forma do SN”, “expressão do sujeito do infinitivo” e “forma do item qualificador”. Neste último grupo de fatores realizou-se o agrupamento de todas as categorias que representavam formas nominais formadas por sufixação em um único fator, que se chamou de “formas derivadas”.

As frequências apresentadas na tabela 11 indicam que o uso da preposição *de* é alto quando o objeto do verbo infinitivo é codificado como sintagma definido, assim como nos casos em que o mesmo é apagado: 1 em cada 4 ocorrências de elipse do SN co-ocorre com a preposição *de*. Em relação a “expressão do sujeito do infinitivo”, as ocorrências com sujeito expreso predominaram nas sentenças interligadas pelo conector *para*: a cada sete ocorrências de *para* apenas uma delas não está acompanhada pelo sujeito do verbo no infinitivo (1:7).

**Tabela 11**  
**Variável forma de conexão entre as orações na amostra de escrita (NILC)**

		FORMA DE CONEXÃO ENTRE AS ORAÇÕES							
		Ausência de conector		Preposição <i>de</i>		Preposição <i>para</i>		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma do SN (objeto do verbo infinitivo)	elipse	2	0,8%	9	25%	0	0%	11	3,5%
	SN definido	135	51,3%	17	47,2%	7	50%	159	50,8%
	SN genérico	48	18,3%	2	5,6%	2	14,3%	52	16,6%
	SN indefinido	61	23,2%	2	5,6%	3	21,4%	66	21,1%
	pronome pessoal	14	5,3%	2	5,6%	1	7,1%	17	5,4%
	resumitivo/fórico	3	1,1%	4	11,1%	1	7,1%	8	2,6%
Exp. sujeito do infinitivo	não expreso	261	99,2%	35	97,2%	2	14,3%	298	95,2%
	expreso	2	0,8%	1	2,8%	12	85,7%	15	4,8%
Forma do elemento nominal	adjs primitivos	120	45,6%	31	86,1%	9	64,3%	160	51,1%
	formas derivadas	143	54,4%	5	13,9%	5	35,7%	153	48,9%
Total		26	100%	36	100%	14	100%	313	100%

Na distribuição de acordo com tipo de item qualificador, as aqui chamadas “formas derivadas” foram um pouco mais recorrentes quando nenhum conector foi utilizado entre as

orações. Ao passo que os adjetivos primitivos predominaram nos contextos em que houve a utilização de uma das preposições. Ainda nos contextos em que se registrou o uso de conector, os elementos nominais predominantes foram adjetivos do tais como difícil, fácil, bom e ruim (denominados na tabela “adjetivos primitivos”).

#### 4.1.2.1 Forma de conexão e forma do SN na escrita

Com a intenção de averiguar uma possível correlação entre o uso do conector *de* e a ocorrência do apagamento do objeto do infinitivo, realizou-se o cruzamento destas variáveis. Segundo a hipótese de trabalho, a preposição *de* deve ser mais recorrente nos contextos em que o verbo infinitivo está desacompanhado de seu complemento, seja por haver sido alçado ou por ter sido apagado, como em (38).

- (38) OS CONSOLES. Nos dois, a conexão na TV é simples, mas o encaixe dos cartuchos é ruim: no Mega (acima), é fácil de colocar a fita e difícil de tirar Ø ; no Nintendo (abaixo), ao contrário. Joystick: o formato Nintendo se adapta melhor à mão, é leve, mas as teclas (sete), longe entre si, dificultam o manuseio. O do Mega, menor e mais bonito, tem cinco teclas juntas: é fácil de apertar Ø (NILC, re94ja23).

A associação obtida no cruzamento da "forma de conexão entre as orações" com a "forma do SN", ficou entre fraca e moderada (quadro 9).

**Quadro 9**

**Testes de significância e associação – Forma de conexão vs. Forma do SN (escrita)**

<i>Likelihood Ratio</i>	48,772	sig 0,000
Tau de Goodman e Kruskal	0,17	sig 0,000
V de Cramer	0,35	sig 0,000

Os pesos relativos deste cruzamento indicam que contextos de sintagmas nominais indefinidos e genéricos inibem o uso da preposição *de* (tabela 12).

**Tabela 12**

**Preposição *de* vs. Forma do SN – pesos relativos da amostra de escrita**

		PREPOSIÇÃO <i>DE</i>			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma do SN	SN indefinido	2	5,5%	0.24	66	21%
	SN genérico	2	5,5%	0.33	52	16,7%
	SN definido	17	47,3%	0.57	159	50,8%
	pronome pessoal	2	5,5%	0.52	17	5,5%
	resumitivo/fórico	4	11,2%	0.94	8	2,5%
	elipse	9	25%	0.98	11	3,5%
Total		36	100%		313	100%

Input 0,059 Log likelihood -75,146 Sig. 0,000

Em contrapartida, pronomes resumitivos ou fóricos, bem como o apagamento (elipse do SN) são os contextos mais favoráveis para o uso desta preposição, corroborando, então, a hipótese de pesquisa. Exemplo:

- (39) Tenho formação animal, gosto de cheirar, sentir, tocar, mexer. Posso até admirar a construção intelectual mas só vale se me comover. É como o amor por uma mulher, a gente pensa que gostou do nariz, do cabelo, dos olhos, mas é mentira - foi uma emoção e **isso** é difícil **de** explicar (NILC, id96mr16).

#### 4.1.2.3 Forma do item qualificador na escrita

Em relação ao item qualificador, também não foram obtidos valores altos para a associação entre esta variável independente e a forma de conexão (quadro 10), indicando a existência de uma associação, mas fraca.

**Quadro 10**

**Testes de significância e associação – Forma de conexão vs. Forma do item qualificador (escrita)**

<i>Likelihood Ratio</i>	23,911	sig 0,000
Tau de Goodman e Kruskal	0,05	sig 0,000
V de Cramer	0,26	sig 0,000

Já os valores de pesos relativos indicam que contextos nos quais o item qualificador é um adjetivo do tipo primitivo favorecem fortemente o uso da preposição *de*. Os demais contextos desfavorecem o uso desta variante (tabela 13).

**Tabela 13**

**Preposição *de* vs. Forma do item qualificador – pesos relativos da amostra de escrita**

		PREPOSIÇÃO <i>DE</i>			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma do elemento nominal	formas primitivas	31	86,2%	0.73	160	51,1%
	formas derivadas	5	13,8%	0.26	153	48,9%
Total		36	100		313	100%

**Input 0,045 Log likelihood -75,579 Sig. 0,000**

Os exemplos (40) e (41) ilustram o uso do conector *de* junto de adjetivos classificados neste trabalho como “formas primitivas”.

- (40) Atlas Folha é **fácil de** usar (NILC, br94ag10).

- (41) O porquê é **difícil de** explicar (NILC, re94ma15).

Os exemplos de (42) a (44) ilustram o uso das “formas derivadas”, formadas, respectivamente, pelos morfemas *-vel* e *-ante* e *-ado*.

(42) É **impossível** Ø detectar quarks isoladamente (NILC, br94ab27).

(43) É **importante** Ø ter um mapa atualizado (NILC, br94ag15).

(44) É **complicado** Ø justificar isso para a sociedade (NILC, br94no19).

Estes resultados foram coerentes com aqueles obtidos nos estudos preliminares (NEVES, 2003; 2004; MITTMANN, 2003; MAY, 2004).

#### 4.1.2.4 Forma de conexão e Expressão do sujeito do infinitivo na escrita

A expressão do sujeito do verbo infinitivo aparenta ser fortemente relacionada ao uso da preposição *para*. Tomando este conector como valor de aplicação da regra na análise multidimensional do GoldVarb, este é o único grupo de fatores selecionado (tabela 14).

O cruzamento entre a “forma de conexão” e a “expressão de sujeito do infinitivo” apresentou a maior diferença nos resultados dos testes de associação aplicados. A Tau de Goodman e Kruskal indica uma associação moderada, já o V de Cramer uma associação muito forte.

**Tabela 14**  
**Preposição *para* vs. Expressão do sujeito do inf. – pesos relativos da amostra de escrita**

		PREPOSIÇÃO PARA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Exp. sujeito do infinitivo	não expresso	2	14,3%	0.42	298	95,2%
	expresso	12	85,7%	0.99	15	4,8%
Total		14	100%		313	100%

Input 0,009 Log likelihood -19,507 Sig. 0,000

Como a Tau de Goodman e Kruskal utiliza a PRE, é provável que as grandes diferenças nos valores observados nas marginais das variável dependente e variável independente tenham criado uma maior possibilidade de predições incorretas, fazendo o valor desta medida tender mais para baixo.

**Quadro 11**  
**Testes de significância e associação - Forma de conexão vs. Exp. sujeito do infinitivo (escrita)**

Likelihood Ratio	76,291	sig 0,000
Tau de Goodman e Kruskal	0,19	sig 0,000
V de Cramer	0,82	sig 0,000

A relação observada entre o conector *para* e a presença de sujeito nas sentenças complementa os resultados obtidos no tocante à ordem preferencial dos constituintes nestas

sentenças. Nas sentenças em que o verbo infinitivo tem seu sujeito expresso, este será introduzido pela preposição *para*, e a ordem da sentença será a canônica, o que significa dizer que o verbo infinitivo será seguido de seu complemento. Exemplos:

(45) Nesse momento, é mais importante **para nós** dar um padrão tático ao time (NILC, sp96fe07).

(46) [...] é crucial **para o país** concluir a liberalização da economia, atrair investimentos e estabilizar os preços (NILC, br94ab17).

(47) Até agora está difícil **para ele** contabilizar os dividendos políticos adquiridos com tal empreitada (NILC, op94ag14).

As hipóteses de trabalho concernentes ao uso de *para* estabeleciam que este conector deveria predominar nos contextos nos quais o verbo infinitivo é acompanhado de seu sujeito, casos em que a oração combinada preserva as características de oração substantiva.

Os dados analisados corroboram com esta hipótese, e a interpretação proposta para tal fenômeno é a de que a preposição *para* tem a função de síndeto em uma relação hipotática entre as orações, funcionando ao mesmo tempo como conector e atribuidor de caso, função já descrita por Rocha (1989).

Retomando Lehmann (1988), observamos que a presença de ambos, sujeito e complemento do verbo infinitivo, implica um baixo grau de entrelaçamento entre as orações, já que há menos constituintes compartilhados entre a matriz e a marginal, o verbo da marginal não apresenta sinais de nominalização, apresentando, portanto, uma baixa dessentencialização.

## 4.2 Resultados da amostra de fala da região sul (VAR SUL)

Na amostra de fala, além das variáveis lingüísticas, também foram controladas as variáveis extralingüísticas “sexo”, “faixa etária”, “nível de escolaridade” e “cidade de origem” do informante. As distribuições dos cruzamentos entre as variáveis dependentes “configuração da sentença” e “forma de conexão” estão divididas em duas tabelas cada uma, uma mostrando as variáveis lingüísticas que se mostraram significativas para o fenômeno em estudo, e outra exibindo a distribuição das ocorrências para as variáveis extralingüísticas. Entre estas últimas, nenhuma atingiu a significância mínima nos testes do

programa SPSS 10.0, e apenas a escolaridade foi selecionada em uma das rodadas estatísticas do GoldVarb 2001, em cruzamento com a “forma de conexão”.

Decidiu-se exibir as frequências de acordo com a estratificação social completa, a título de descrição da amostra, para cada uma das variáveis dependentes. As ocorrências codificadas como “elipse do SN” na variável “configuração da sentença” foram marcadas com o valor “não se aplica” para a realização da análise estatística do programa GoldVarb 2001.

Para eliminar as células com valores iguais a zero também foram desconsideradas as categorias “SN indefinido” e “elipse” na variável independente “forma do SN”. Por este motivo, nas tabelas em que a variável “configuração da sentença” está envolvida, o número total de ocorrências é 63. Nas demais tabelas o total de ocorrências permanece igual a 89.

#### 4.2.1 Configuração da sentença na fala da região sul (VARSUL)

A distribuição das ocorrências da variável “configuração da sentença” em cruzamento com os grupos de fatores extralingüísticos (estratificação social dos informantes) está apresentada na tabela 15. Neste recorte da amostra de fala não estão sendo considerados 12 informantes da faixa etária dos 15 aos 24 anos da cidade de Florianópolis.

**Tabela 15**  
**Variável configuração da sentença na amostra de fala da Região Sul – fatores sociais**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canônica		Oração infinitiva deslocada		Elipse do SN		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Sexo	feminino	5	45,5%	23	57,5%	5/12	41,7%	13	50%	46	51,7%
	masculino	6	54,5%	17	42,5%	7/12	58,3%	13	50%	43	48,3%
Faixa etária	entre 25 e 49	9	81,8%	31	77,5%	10/12	83,3%	14	53,8%	64	71,9%
	mais de 50	2	18,2%	9	22,5%	2/12	16,7%	12	46,2%	25	28,1%
Nível de escolaridade	de 1-4 anos	4	36,4%	17	42,5%	5/12	41,7%	10	38,5%	36	40,4%
	de 5-8 anos	6	54,5%	13	32,5%	6/12	50%	11	42,3%	36	40,4%
	de 9-11 anos	1	9,1%	10	25%	1/12	8,3%	5	19,2%	17	19,1%
Cidade de origem	Curitiba	6	54,5%	17	42,5%	5/12	41,7%	12	46,2%	40	44,9%
	Florianópolis	1	9,1%	8	20%	4/12	33,3%	4	15,4%	17	19,1%
	Porto Alegre	4	36,4%	15	37,5%	3/12	25%	10	38,5%	32	36%
Total		11	100%	40	100%	12	100%	26	100%	89	100%

O total de ocorrências coletadas neste corpus foi 89. As construções de alçamento somam um total de 11 ocorrências (12,3%), na ordem canônica reuniu-se 40 casos (45%), os casos de oração infinitiva deslocada são em número de 12 (13,5%) e os casos de apagamento do sintagma nominal somam um total de 26 ocorrências (29,2%). Homens e mulheres mostraram um comportamento muito parecido no uso de todas as formas. Em relação ao alçamento, a faixa etária que mais produziu esta configuração foi a mais jovem da amostra, situada entre 25 e 49 anos. Curitiba foi a localidade onde as construções alçadas foram mais utilizadas.

A ordem canônica reuniu praticamente metade das ocorrências, e foi a ordenação mais utilizada em quase todos os níveis da estratificação, à exceção da faixa etária de mais de 50 anos, nesta faixa etária o valor modal de configuração da sentença está na categoria “elipse do SN”.

A tabela 16 apresenta as frequências para os cruzamentos entre a configuração da sentença e os grupos de fatores “forma de conexão”, “forma do SN”, “distância referencial” e “status informacional”, estes são os grupos que obtiveram nível de significância mínimo nos testes do programa SPSS 10.0.

**Tabela 16**  
**Variável configuração da sentença na amostra de fala da Região Sul – fatores linguísticos**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canônica		Oração infinitiva desl.		Elipse do SN		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma de conexão entre as orações	ausência	3	27,3%	30	75%	4/12	33,3%	4	15,4%	41	46,1%
	preposição <i>de</i>	7	63,6%	5	12%	0/12	0%	16	61,5%	28	31,5%
	preposição <i>para</i>	1	9,1%	5	12%	8/12	66,7%	6	23,1%	20	22,5%
Forma do SN (objeto do verbo infinito)	elipse	0	0%	1	2,5%	2/12	16,7%	25	96,2%	28	31,5%
	SN definido	6	54,5%	16	40%	2/12	16,7%	0	0%	24	27%
	SN genérico	3	27,3%	7	17,5%	2/12	16,7%	0	0%	12	13,5%
	SN indefinido	0	0%	10	25%	5/12	41,7%	1	3,8%	16	18%
	pronome pessoal	1	9,1%	3	7,5%	1/12	8,3%	0	0%	5	5,6%
	resumitivo/fórico	1	9,1%	3	7,5%	0/12	0%	0	0%	4	4,5%
Exp. do sujeito do infinitivo	não expresso	10	90,9%	22	55%	11/12	91,7%	20	76,9%	63	70,8%
	expresso	1	9,1%	18	45%	1/12	8,3%	6	23,1%	26	29,2%
Distância da última menção	primeira menção	3	27,3%	14	35%	1/12	8,3%	0	0%	18	20,2%
	1 a 3 orações	5	45,5%	20	50%	5/12	66,7%	22	84,6%	55	61,8%
	4 a 10 orações	3	27,3%	6	15%	3/12	25%	4	15,4%	16	18%
Status informacional	dado	8	72,7%	26	65%	11/12	91,7%	26	100%	71	79,8%
	inferível	4	18,2%	1	2,5%	1/12	8,3%	0	0%	4	4,5%
	novo	1	9,1%	13	32,5%	0/12	0%	0	0%	14	15,7%
Total		11	100%	40	100%	12	100%	26	100%	89	100%

A configuração canônica da sentença ocorreu em 75% dos casos sem que houvesse um conector ligando as orações encaixadas. Quando a oração encaixada é deslocada para a esquerda do verbo finito, esta é geralmente introduzida pela preposição *para* (66,7% das vezes), e nunca com a preposição *de*. Já quando há apagamento do objeto do verbo infinitivo, há um aumento no uso da preposição *de* e uma diminuição no uso da justaposição.

#### 4.2.1.1 Alçamento e forma de conexão entre as orações na fala

Os resultados desta amostra foram muito semelhantes aos da escrita. A maior parte das construções de alçamento na fala foi realizada concomitantemente à preposição *de* (63,6%). A associação entre as variáveis “configuração da sentença” e “forma de conexão” (quadro 8) é entre moderada e forte na amostra de fala.

**Quadro 12**

**Testes de significância e associação – Configuração da sentença vs. Forma de conexão (fala)**

<i>Likelihood Ratio</i>	46,880	sig 0,000
Tau de Goodman e Kruskal	0,21	sig 0,000
V de Cramer	0,51	sig 0,000

A “forma de conexão” (juntamente com o grupo de fatores “expressão do sujeito do infinitivo”) foi selecionada pelo GoldVarb 2001 na aplicação da regra do alçamento. Os pesos relativos para “forma de conexão entre as orações” na aplicação da regra do alçamento estão apresentados na tabela 17.

**Tabela 17**

**Alçamento vs. Forma de conexão – pesos relativos da amostra de fala**

		ALÇAMENTO		Total da amostra		
		Freq.	%	P.R.	Freq.	%
Forma de conexão entre as orações	preposição <i>para</i>	1	9,1%	0.30	14	22,3%
	ausência	3	27,3%	0.37	37	58,7%
	preposição <i>de</i>	7	63,6%	0.93	12	19%
Total		11	100%		63	100%

**Input 0,094 Log likelihood -19,335 Sig. 0,018**

Os valores de pesos relativos indicando um favorecimento do alçamento com a preposição *de*, e a inibição do mesmo com a justaposição e a preposição *para*. Exemplos:

(48) Bah! esse psicotécnico deve ser horrível de fazer né? (VARSUL, POA 05).

(49) Depois está difícil também pra dar educação, estudo (VARSUL, POA 12).



(50) Mas está tão difícil ~~o~~ comprar um terreno , fazer casa (VARSUL, CTB 08).

A hipótese de que a preposição *de* seria predominante nos contextos de alçamento do sintagma nominal foi confirmada, entretanto novamente foi observada baixa frequência das construções alçadas, bem como do uso deste conector.

Os informantes mais jovens utilizaram com mais frequência do que os mais velhos as estruturas alçadas, entretanto, os mais jovens utilizaram todas as variantes com frequência superior aos mais velhos, sem que fosse encontrada associação entre a “configuração da sentença” e os grupos de fatores sociais.

#### 4.2.1.2 Forma do SN na fala

O comportamento dos dados para o grupo de fatores “forma do SN” está bastante semelhante ao observado na amostra de textos escritos. Os sintagmas alçados a sujeito foram na maioria definidos (54% dos SNs alçados), e não foi registrado o uso de sintagmas nominais indefinidos na posição alçada. Os resultados dos testes de associação entre a “configuração da sentença” e “forma do SN” indicam associação mais forte entre estas variáveis na amostra de fala do que na de escrita. Além do valor mais alto para a Tau de Goodman e Kruskal, a diferença entre este e o valor de V de Cramer é menor (quadro 13).

**Quadro 13**

**Testes de significância e associação – Configuração da sentença vs. Forma do SN (fala)**

<i>Likelihood Ratio</i>	96,008	sig 0,000
Tau de Goodman e Kruskal	0,40	sig 0,000
V de Cramer	0,56	sig 0,000

O grupo de fatores “forma do SN” não foi selecionado pelo programa GoldVarb 2001 como relevante para a variável “configuração da sentença”. Por este motivo, em relação à hipótese de trabalho para esta variável, não é possível obter um resultado conclusivo, no sentido de saber se as construções de alçamento “selecionam” referentes definidos (50) e genéricos (51), já que os testes de associação efetuados se baseiam apenas nos valores das marginais da tabela.

(51) Mas o camarão mesmo, realmente, está caríssimo, né? E o peixe , inclusive, aqui no Estreito, é difícil de se encontrar. (VARSUL, FLP 11)

(52) Olha, **churrasco** é bem fácil de preparar, né ? existem várias maneiras (VARUSL, POA 04).

#### 4.2.1.3 Alçamento e expressão do sujeito do infinitivo na fala

O alçamento ocorreu 90,9% das vezes em orações sem sujeito do infinitivo expresso. A associação da variável “configuração da sentença” e o grupo de fatores “expressão do sujeito do infinitivo” mostrou valores bem afastados nas duas medidas (quadro 14).

Para este cruzamento as análises realizadas pelos programas SPSS e GoldVarb 2001 divergiram bastante. O GoldVarb 2001 selecionou o grupo de fatores “expressão do sujeito do infinitivo”, que obteve associação fraca, e deixou de fora o grupo “forma do SN”, que mostrou uma associação mais forte.

**Quadro 14**

**Testes de significância e associação – Configuração da sentença vs. Exp. do sujeito do infinitivo (fala)**

<i>Likelihood Ratio</i>	10,794	sig 0,000
Tau de Goodman e Kruskal	0,05	sig 0,000
V de Cramer	0,34	sig 0,000

O número total de ocorrências não é muito grande, e as ocorrências não estão distribuídas uniformemente entre as variáveis, então os resultados obtidos nos testes de associação e de pesos relativos (exibidos na tabela 18) devem ser observados com cautela.

**Tabela 18**

**Alçamento vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala**

		ALÇAMENTO		Total da amostra	
		Freq.	%	Freq.	%
Exp. sujeito de infinitivo	expresso	1	9,1%	20	31,8%
	não expresso	10	90,9%	43	68,2%
Total		11	100%	63	100%

**Input 0,094 Log likelihood -19,335 Sig. 0,018**

A “expressão do sujeito do infinitivo” não obteve resultados significativos na amostra de escrita para o alçamento, mas obteve para a “ordem canônica”.

Os resultados exibidos aqui não contrariam os resultados obtidos para a escrita, já que a presença de um sujeito expresso inibe o alçamento, que são mais comuns quando a ordem é a canônica. Exemplo:

(53) Antes de quinze dias era difícil **você** conseguir um ortopedista (VARUSL, CTB 19).

#### 4.2.1.4 Distância referencial e status informacional na fala

Estas variáveis foram propostas com o intuito de captar a topicalidade do referente passível de alçamento. A hipótese de trabalho em relação a estas variáveis era de que acessibilidade mais alta (menor distância referencial) e o status informacional “dado” ou “inferível” seriam fatores que contribuiriam para o alçamento, exemplificado em (53). Em contrapartida, baixa acessibilidade (maior distância referencial) e status informacional “novo” deveriam inibir o alçamento.

(54) E: **Fórmula Um**, passa na televisão, ou vôlei, assistes essas coisa?

F: **Fórmula Um** hoje é difícil ver, né? Antes eu via. Quando o Senna corria a gente via bastante, né? (VARSL, FLP J02).

Os resultados para estas variáveis não são conclusivos, tendo em vista que nenhuma delas foi selecionada pelo programa GoldVarb 2001. O quadro 15 mostra os valores para a associação entre este grupo e a “configuração da sentença”, que apresentou associação fraca. O único contexto de acessibilidade bem definido foi nos casos de elipse do sintagma nominal, que nunca ocorrem quando a distância referencial é grande.

**Quadro 15**

**Testes de significância e associação – Configuração da sentença vs. Distância referencial (fala)**

<i>Likelihood Ratio</i>	19,932	sig 0,003
Tau de Goodman e Kruskal	0,08	sig 0,000
V de Cramer	0,29	sig 0,000

Pelas frequências, verifica-se que as construções em estudo parecem ser típicas de contextos discursivos onde o objeto do verbo infinitivo é uma informação dada no discurso precedente (79,8% das ocorrências)<sup>23</sup>. Dentro deste conjunto de dados, quando ocorre a introdução de referentes novos no discurso nestas construções, a configuração preferencial é a ordem canônica, o que obedece aos princípios icônicos gerais de espacialização e sequência (GIVÓN 2001b).

No quadro 16 são apresentados os valores de associação para o grupo de fatores “status informacional”. Aqui também foram observadas diferenças muito grandes entre os valores dos dois testes de associação realizados.

<sup>23</sup> As construções estudadas aqui apresentam obrigatoriamente um e no máximo dois referentes, ocupando as funções sintáticas de objeto do verbo infinito e sujeito deste mesmo verbo.

Quadro 16

Testes de significância e associação – Configuração da sentença vs. Status informacional (fala)

<i>Likelihood Ratio</i>	26,098	sig 0,000
Tau de Goodman e Kruskal	0,11	sig 0,000
V de Cramer	0,36	sig 0,000

É possível que a distância referencial e o status informacional atuem indiretamente, e que outros fatores sejam os decisivos na opção pelo alçamento. No exemplo (54), é provável que a presença de um interferente em potencial (vôlei) aliada à alta acessibilidade tenha sido a razão para o uso da estrutura alçada, já que o apagamento do item geraria dúvida na interpretação da resposta (54a):

- (54a) E: Fórmula Um, passa na televisão, ou vôlei, assistes essas coisas?  
F: Ø hoje é difícil ver, né? Antes eu via.

A ambigüidade só seria esclarecida na próxima sentença (53b), pela introdução do referente Senna e do verbo correr:

- (54b) Quando o Senna corria a gente via bastante, né?

Em outros casos, é possível que também a persistência seja um fator decisivo, ainda que esta categoria não tenha se mostrado significativa em nenhuma das amostras. No exemplo (55), além dos interferentes potenciais (cocada, brinquedo, coisa pra vender), há enorme recorrência do referente alçado.

- (55) F: Tem pinhão tem tudo, na Festa do Pinhão dizem que tem tudo, né? Pizza de pinhão, é cachorro-quente de pinhão, bolo de pinhão, tudo de pinhão. Assim que é festa, né? porque aqui a Festa da **Laranja**. O que que às vezes você encontra? Só cocada.  
E: Só cocada?  
F: Está cheio de cocada. **Laranja** é difícil tu encontrares na festa ali. E: Engraçado, né?  
F: É, brinquedo tem, coisa pra vender tem, mas o que encontra é cocada. Não tem nada assim, doce de **laranja**, existe bastante coisa assim de **laranja**, bolo tem de laranja, né? Sobremesa tem de **laranja**, mas eles não fazem. Só **laranja** assim (VARSL, FLP J06).

Neste caso, o alçamento pode ser interpretado como marcação da importância do referente no discurso seguinte, dando ênfase para o tópico, o assunto sobre o que o falante deseja comentar.

Avalia-se que talvez fossem obtidos resultados mais satisfatórios se a topicalidade fosse considerada uma variável independente contínua complexa, na qual a “distância referencial”, o “status informacional” e a “persistência” fossem critérios que servissem para medir o grau da topicalidade, ao invés de serem tratadas como categorias discretas independentes.

#### 4.2.1.5 Ordem canônica e forma de conexão na fala

Como muitas ocorrências se concentram na variante “ordem canônica” da configuração da sentença, foram realizadas novas rodadas estatísticas no GoldVarb 2001 considerando-se a ordem canônica como aplicação da regra em oposição às demais configurações. A tabela 19 mostra os pesos relativos para a aplicação desta regra.

**Tabela 19**

**Ordem canônica vs. Forma de conexão – pesos relativos da amostra de fala**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma de conexão entre as orações	preposição <i>de</i>	5	12,5%	0.11	12	19
	preposição <i>para</i>	5	12,5%	0.18	14	22,3
	ausência	30	75%	0.78	37	58,7
Total		40	100%		63	100%

Input 0,788 Log likelihood -20,984 Sig. 0,039

Novamente, a forma de conexão entre as orações foi um dos grupos de fatores selecionados para a configuração da sentença. A justaposição é o contexto que mais favorece a ordem canônica da sentença, ao passo que as preposições *para* e *de* a desfavorecem. Exemplo:

(56) Antes de quinze dias era difícil você conseguir um ortopedista (VARSUL, CTB 19).

Este resultado complementa aquele obtido nos pesos relativos para o “alçamento” e a “forma de conexão”

#### 4.2.1.6 Ordem canônica e expressão de sujeito do infinitivo

O segundo grupo selecionado na análise multidimensional do GoldVarb 2001 para a regra de uso da ordem canônica foi “expressão de sujeito do infinitivo” (tabela 20).

**Tabela 20**

**Ordem canônica vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Expressão do sujeito do infinitivo	ausente	22	55%	0.25	43	68,2%
	presente	18	45%	0.91	20	31,8%
Total		40	100%		63	100%

Input 0,788 Log likelihood -20,984 Sig. 0,039

O fator que induz fortemente a aplicação desta regra é a o sujeito expresso, ao passo que a ausência do sujeito desfavorece esta regra. Este resultado complementa aquele apresentado anteriormente na tabela 18 (na qual a aplicação da regra era o alçamento).

Exemplo:

- (57) Naquele tempo era muito difícil a gente ver um abacate. Pelo menos, aqui no Caieira, era muito difícil ver um abacate (VARSUL, FLP 12)

A hipótese de trabalho para esta variável levava em consideração que a presença do sujeito do verbo infinitivo é um parâmetro de menor entrelaçamento das sentenças (LEHMANN 1988), assim, a hipótese de que as orações combinadas por justaposição são menos integradas do que aquelas onde ocorre o conector *de* é corroborada.

#### 4.2.1.7 Ordem canônica e status informacional

A tabela 21 apresenta valores de pesos relativos para o grupo de fatores “status informacional”. Complementando os resultados já exibidos, nota-se que referentes novos privilegiam a ordenação canônica da sentença, ou seja, tendem a ser apresentados pospostos ao verbo no infinitivo, ocupando uma posição menos tópica, como exemplificado em (57).

**Tabela 21**  
**Ordem canônica vs. Status informacional – pesos relativos da amostra de fala**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Status informacional	inferível	1	2,5%	0.03	4	6,3%
	dado	26	65%	0.37	45	71,4%
	novo	13	32,5%	0.94	14	22,3%
Total		40	100%		63	100%

Input 0,788 Log likelihood -20,984 Sig. 0,039

- (58) E: Quem ganha nenê, já de primeira viagem, já é duro, né? e gêmeos ainda

F: Gêmeos ainda nós tínhamos pouca experiência, né? tudo, né? Mas tinha minha mãe, né? falecida mãe, né? Ela deu uma grande força, então ela que era a babá, ela que deu os primeiros banhos neles, né? E então

E: Tinha alguém pra ajudar. Choro na madrugada.

F: Eles não choravam de madrugada, porque eles ficavam acordados a noite inteira. Porque eles trocaram a noite pelo dia. O dia inteiro eles dormiam e depois quando era de noite eles não dormiam, ficavam brincando no berço.

E: Desde pequeninhos?

F: É, desde pequeninhos. Aí depois, daí me ensinaram que era bom fazer chá de alface, chá de alfaça, então eu fui atrás e nós carcávamos chá neles, lá.

E: De dia. De noite.

F: Daí não deixava eles dormirem de dia, né? Não deixava eles dormirem. Daí quando era de tardezinha, a mulher aplicava chá nos bichos. Aí de noite eles dormiam.

Este resultado corrobora com a hipótese de trabalho de que referentes novos têm menos chance de serem alçados, estando de acordo com os princípios icônicos de organização do discurso.

#### 4.2.1.8 Ordem canônica e escolaridade na fala

O único grupo de fatores extralingüísticos selecionado em todas as análises realizadas com o programa GoldVarb 2001 foi o nível de escolaridade, considerando-se a ordem canônica como valor de aplicação da regra (tabela 22).

**Tabela 22**  
**Ordem canônica vs. Nível de escolaridade – pesos relativos da amostra de fala**

		ORDEM CANÔNICA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Nível de escolaridade	1 a 4 anos	17	42,5%	0.65	26	41,3%
	5 a 8 anos	13	32,5%	0.22	25	39,7%
	9 a 11 anos	10	25%	0.74	12	19%
Total		40	100%		63	100%

**Input 0,788 Log likelihood -20,984 Sig. 0,039**

Estes valores de pesos relativos não representam uma linha ascendente de uso da configuração canônica à medida em que aumenta o tempo de escolarização do informante. Informantes das faixas mais altas e mais baixas são aqueles que utilizam mais as construções canônicas, ao passo que a faixa de escolarização intermediária (entre 5 e 8 anos de escolaridade) é a que mais utiliza as formas inovadoras (alçamento e oração infinitiva deslocada).

Neste estudo não foram formuladas hipóteses relacionadas à escolarização dos falantes, já que este pacote de variação não apresenta formas socialmente estigmatizadas. Este resultado é, portanto, intrigante. Seria razoável supor que falantes mais escolarizados pudessem preferir a configuração canônica por terem um maior contato com a escrita, modalidade em que a variação no uso das diferentes configurações é menor, com predominância da forma canônica. Entretanto, o mesmo comportamento é observado nos informantes de escolaridade mais baixa, o que refuta esta hipótese. Vale salientar aqui que,

na tabulação cruzada realizada através do SPSS 10.0, não foi detectada associação entre a configuração da sentença e o nível de escolaridade, assim, é possível que este efeito esteja sendo causado pela interação da escolaridade com outras variáveis, ou por conta do número baixo de informantes por célula.

#### 4.2.1.9 Deslocamento da oração infinitiva na fala

Tomando o deslocamento de toda a oração infinitiva como aplicação da regra na análise multidimensional, os grupos de fatores relevantes são a “forma de conexão entre as orações” (tabelas 23)<sup>24</sup> e a “expressão de sujeito do infinitivo” (tabela 24).

**Tabela 23**

**Oração infinitiva deslocada vs. Forma de conexão – pesos relativos da amostra de fala**

		ORAÇÃO INFINT. DESL.			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma de conexão entre as orações	ausência	4	33,3%	0.34	37	72,5%
	preposição <i>para</i>	8	66,7%	0.85	14	27,5%
Total		12	100%		51	100%

**Input 0,151 Log likelihood -21,309 Sig. 0,009**

Como esperado, o uso da preposição *para* está fortemente relacionado ao deslocamento da oração infinitiva inteira para a posição inicial da sentença. Isto traz evidências para a hipótese de que a preposição *para* é utilizada para introduzir um constituinte oracional. A possibilidade das orações introduzidas por *para* aparecem tanto antepostas quanto pospostas em relação à oração matriz indica que o estatuto das construções com *para* é hipotático, e não de encaixamento.

O deslocamento da oração encaixada inteira para a posição inicial na sentença também é favorecido quando sujeito do verbo infinitivo é omitido (tabela 24)<sup>25</sup>. Ainda que este grupo de fatores tenha sido selecionada pelo programa GoldVarb 2001, a convergência na análise multidimensional não foi atingida até 20ª interação.

24 O total de ocorrências exibido na tabela 20 é 51 por terem sido eliminadas as ocorrências com a preposição *de*, já que não foi coletada nenhuma ocorrência deste conector em oração infinitiva deslocada.

25 Quando o sujeito é expreso, é mais provável que a ordenação da sentença seja a canônica (como exibido na tabela 20).



**Tabela 24****Oração infinitiva deslocada vs. Exp. do sujeito do infinitivo – pesos relativos da amostra de fala**

		ORAÇÃO INFINT. DESL.			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Expressão do sujeito do infinitivo	expresso	1	8,4%	0.22	20	68,2%
	não expresso	11	91,6%	0.64	43	31,8%
Total		12	100%		63	100%

**Input 0,151 Log likelihood -21,309 Sig. 0,009**

No nível 1 da análise multidimensional, quando as variáveis independentes são testadas uma a uma em relação à variável dependente, tanto a rodada para a “forma de conexão” quanto para a “expressão do sujeito” atingiram a convergência na quinta interação. Quando a convergência não é atingida, ou demora muito para ser atingida, implica em uma confiabilidade menor nos resultados exibidos.

#### 4.2.2 Forma de conexão entre as orações na fala da região Sul (VARSUL)

A tabela 25 apresenta a distribuição das ocorrências da variável dependente "forma de conexão entre as orações" em cruzamento com os grupos de fatores extralingüísticos.

**Tabela 25****Variável forma de conexão entre as orações na amostra de fala da Região Sul – fatores sociais**

		FORMA DE CONEXÃO ENTRE AS ORAÇÕES							
		Ausência de conector		Preposição <i>de</i>		Preposição <i>para</i>		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Sexo	feminino	20	48,8%	14	50%	12	60%	46	51,7%
	masculino	21	51,2%	14	50%	8	40%	43	48,3%
Faixa etária	entre 25 e 49	33	80,5%	17	60,7%	14	70%	64	71,9%
	mais de 50	8	19,5%	11	39,3%	6	30%	25	28,1%
Nível de escolaridade	de 1-4 anos	17	41,5%	9	32,1%	10	50%	36	40,4%
	de 5-8 anos	17	41,5%	11	39,3%	8	40%	36	40,4%
	de 9-11 anos	7	17,1%	8	28,6%	2	10%	17	19,1%
Cidade de origem	Curitiba	21	51,2%	11	39,3%	8	40%	40	44,9%
	Florianópolis	8	19,5%	5	17,9%	4	20%	17	19,1%
	Porto Alegre	12	29,3%	12	42,9%	8	40%	32	36%
Total		41	100%	28	100%	20	100%	89	100%

Neste cruzamento, nenhum destes grupos de fatores foi selecionado na análise multidimensional do programa GoldVarb 2001, ou atingiu a significância mínima de 0,05 nos testes do SPSS 10.0. Estes resultados são exibidos aqui a caráter de descrição. As ocorrências de acordo com a forma de conexão entre as orações se dividem em 41 casos de justaposição

(46%), 28 casos de conexão através da preposição *de* (31,5%) e 20 de conexão através da preposição *para* (22,5%).

Novamente, mulheres e homens apresentaram um comportamento lingüístico muito semelhante, à exceção do uso da preposição *para*, que foi mais utilizada pelas mulheres. A maior parte das ocorrências da sentença em estudo foi produzida pelos informantes da faixa etária dos 25 até os 49 anos. A justaposição foi a forma de conexão entre as orações mais utilizada por esta faixa etária.

A preposição *de* foi a forma de conexão mais utilizada por aqueles com maior nível de escolaridade (9 a 11 anos), e a justaposição foi a forma de conexão preferida pelos informantes com menos de 8 anos de escolaridade. De modo geral, o conector menos utilizado foi a preposição *para*, sendo preferido pelos mais jovens, o que pode ser um indício de que esta forma, ainda que já tenha sido registrada no português antigo, é uma forma inovadora, no sentido de que há um aumento no seu uso pelos mais jovens. Foi uma surpresa a alta taxa de uso de construções com *de* entre os mais velhos, pois o esperado era que a faixa etária mais velha utilizasse mais freqüentemente a estrutura mais bem estabelecida, em que a combinação entre as orações se dá por justaposição (ausência de conector).

As freqüências para os grupos de fatores lingüísticos estão apresentados na tabela 26.

**Tabela 26**  
**Variável forma de conexão entre as orações na amostra de fala da Região Sul – fatores lingüísticos**

		FORMA DE CONEXÃO ENTRE AS ORAÇÕES							
		Ausência de conector		Preposição <i>de</i>		Preposição <i>para</i>		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma do SN (objeto do verbo infinito)	elipse	5	12,2%	16	57,1%	7	35%	28	31,5%
	SN definido	16	39%	4	14,3%	4	20%	24	27%
	SN genérico	6	14,6%	3	10,7%	3	15%	12	13,5%
	SN indefinido	10	24,4%	1	3,6%	5	25%	16	18%
	pronome pessoal	2	4,9%	2	7,1%	1	5%	5	5,6%
	resumitivo/fórico	2	4,9%	2	7,1%	0	0%	4	4,5%
Expressão do sujeito do infinitivo	não expresso	28	68,3%	20	71,4%	15	75%	63	70,8%
	expresso	13	31,7%	8	28,6%	5	25%	26	29,2%
Presença de materiais interven.	ausência	36	87,8%	23	82,1%	12	60%	71	79,8%
	presença	5	12,2%	5	17,9%	8	40%	18	20,2%
Total		41	100%	28	100%	20	100%	89	100%

Apenas dois grupos obtiveram a significância mínima nos testes do SPSS 10.0 no cruzamento com a variável dependente "forma de conexão": a "forma do SN" e a "presença de materiais intervenientes". A análise multidimensional do programa GoldVarb 2001 também

selecionou o grupo de fatores "expressão do sujeito do infinitivo" como relevante na aplicação da regra do uso da preposição *para*. Da mesma forma que na amostra de textos escritos, a justaposição foi a variante mais recorrente na amostra de fala. Os valores totais de ocorrências de cada uma das variantes são muito mais próximos entre si (41 ocorrências de justaposição, 28 ocorrências de *de* e 20 ocorrências de *para*) do que os observados na amostra de escrita.

A forma de codificação mais comum do objeto do verbo no infinitivo é a elipse, e esta é preferencialmente acompanhada da preposição *de*. Já a preposição *para* é freqüentemente utilizada quando há um sujeito expreso acompanhando o verbo no infinitivo, do mesmo modo que na amostra escrita.

Esperava-se registrar um número maior de ocorrências com materiais intervenientes, entretanto, o controle de tais elementos não foi rigoroso, no sentido de não serem controladas pausas, repetições e reformulações como material lingüístico interveniente.

#### 4.2.2.1 Preposição *de* na fala

O quadro 17 mostra os valores da associação para o cruzamento da variável dependente "Forma de conexão" e para a "Forma do SN". A associação entre estas variáveis se mostrou moderada.

**Quadro 17**

**Testes de significância e associação – Forma de conexão vs. Forma do SN (fala)**

<i>Likelihood Ratio</i>	24,221	sig 0,007
Tau de Goodman e Kruskal	0,14	sig 0,007
V de Cramer	0,35	sig 0,018

Definindo a preposição *de* como valor de aplicação da regra para o cálculo dos pesos relativos, o grupo de fatores "forma do SN" foi o único selecionado<sup>26</sup> (tabela 27).

26 A variável "configuração da sentença" não foi incluída na análise multidimensional do GoldVarb 2001 como variável dependente para a "forma de conexão entre as orações", já que a associação entre estas variáveis já foi confirmada nos testes anteriores.

**Tabela 27**  
**Preposição de vs. Forma do SN – pesos relativos da amostra de fala**

		PREPOSIÇÃO DE			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma do SN	indefinido	1	3,6%	0.15	16	18%
	definido	4	14,3%	0.35	24	27%
	genérico	3	10,7%	0.47	12	13,5%
	pronome <sup>27</sup>	4	14,3%	0.68	9	10%
	elipse	16	57,1%	0.78	28	31,5%
Total		28	100%		89	100%

Input 0,269 Log likelihood -46,606 Sig. 0,002

Os pesos relativos mostram uma tendência muito forte do uso da preposição *de* quando o SN é apagado na sentença. Exemplos:

(59) porque pra tirar o leite bem cedo, então já dava a ração, que aí a vaca é ficava quietinha, descia bem **o leite**, era fácil, né? **de** tirar **Ø**. (VARSUL, CTB 23).

(60) F: [...] E naquele tempo rádio era besteira, né? Nós usávamos era **galena**, um aparelho que vocês, eu acho que nunca viram.

E: É, eu não conheci.

F: E não conheceram. **É fácil de se fazer Ø**, né? E se liga na luz e funciona uma estação de rádio, né?

Este resultado corrobora com a hipótese de que a recorrência da preposição *de* deveria estar relacionada à estrutura ADJ+DE+INF, em que o verbo infinitivo não é seguido do seu objeto.

Quando o objeto do verbo infinitivo é um pronome, também há o favorecimento do uso a preposição *de*. As duas ocorrências Exemplos:

(61) É, **isso** é difícil, né? **da gente** opinar, porque é Ouro deve ser extraído, é uma riqueza que a gente precisa, (VARSUL, POA 04)

(62) E: Isso assim, o seu pai também trabalha

F: Também trabalha. **É difícil de ver ele aqui**, porque geralmente ele está ali olhe (VARSUL, CTB 05).

#### 4.2.2.2 Preposição *para* na fala

Ao estabelecer a preposição *para* como valor de aplicação da regra, o grupo de fatores selecionado é “expressão do sujeito do infinitivo”(tabela 28), como exemplificado em (62) e

<sup>27</sup> Para a realização das rodadas estatísticas do GoldVarb 2001 foi realizado agrupamento, no grupo "forma do SN", dos fatores "pronome pessoal" e "pronome resumitivo/fórico" em uma mesma categoria "pronomes", pois estes fatores contabilizavam um número muito pequeno de ocorrências, com uma distribuição muito semelhante.

(63).

(63) Então é difícil, sabe? **pra eles** conseguirem contratar [otorrino] (VARSUL, CTB 19).

(64) F: [...] com tudo isso eles ainda cobram por BTN.

E: BTN?

F: BTN, quer dizer, tu começa pagando, vamos supor, mil cruzeiros agora, né? daqui uns dois meses tu já estás pagando dois e meio, quer dizer, é difícil pra ti pagares, principalmente que tem outras coisas pra ti fazeres (VARSUL, POA 12).

**Tabela 28**

**Preposição para vs. Expressão do sujeito do inf. – pesos relativos da amostra de fala**

		PREPOSIÇÃO PARA			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Exp. sujeito de infinitivo	não expresso	15	75%	0.43	63	70,8%
	expresso	5	15%	0.74	26	29,2%
Total		20	100%		89	100%

Input 0,212 Log likelihood -44,622 Sig. 0,018

Ainda que haja um maior número de ocorrências da preposição *para* utilizada em sentenças em que o sujeito do infinitivo foi omitido, os valores dos pesos relativos indicam que a presença do sujeito favorece o uso de *para*. Isto corrobora com a hipótese de que a preposição *para* seria predominar em contextos nos quais o verbo infinitivo é acompanhado de seu sujeito. A presença do sujeito é sinal de menos integração, reforçando a idéia de que sentenças combinadas por *para* são hipotáticas.

#### 4.2.2.3 Justaposição na fala

Por fim, os grupos de fatores selecionados na aplicação da regra de "justaposição entre as orações" foram a "forma do SN" e a "presença de materiais intervenientes". Os pesos relativos da tabela 29 indicam que sintagmas definidos e indefinidos são os que estão mais relacionados à justaposição.

Esperava-se que os sintagmas definidos não favorecessem tão fortemente o uso da variante sem conector, já que estes mesmos referentes deveriam ser predominantes em contextos de alçamento, que por sua vez é contexto preferencial do uso do conector *de*. Ainda assim, os esses pesos relativos complementam os pesos obtidos com a aplicação da regra do uso da preposição *de* (tabela 27).

**Tabela 29**  
**Justaposição vs. Forma do SN – pesos relativos da amostra de fala**

		AUSÊNCIA DE CONECTOR			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Forma do SN	elipse	5	12,2%	0.18	28	31,5%
	pronomes	4	9,8%	0.50	9	10%
	genérico	6	14,6%	0.60	12	13,5%
	indefinido	10	24,4%	0.68	16	18%
	definido	16	39%	0.74	24	27%
Total		41	100%		89	100%

**Input 0,442 Log likelihood -50,384 Sig. 0,013**

Em relação ao grupo de fatores “presença de materiais intervenientes”, o ambiente que mais favorece o uso da variante sem conector é a ausência destes materiais (tabela 30). Os valores dos pesos relativos exibidos nesta tabela são bastante baixos comparados aos obtidos para as demais variáveis.

**Tabela 30**  
**Justaposição vs. Pres. de materiais intervenientes – pesos relativos da amostra de fala**

		AUSÊNCIA DE CONECTOR			Total da amostra	
		Freq.	%	P.R.	Freq.	%
Presença de materiais interven.	presença	5	12,2%	0.23	18	20,2%
	ausência	36	87,8%	0.58	71	79,8%
Total		41	100%		89	100%

**Input 0,442 Log likelihood -50,384 Sig. 0,013**

Para o grupo de fatores "presença de materiais intervenientes", os valores obtidos nos dois testes de associação e também no *Likelihood* são baixos.

**Quadro 18**  
**Testes de significância e associação – Forma de conexão vs. Presença de materiais intervenientes (fala)**

<i>Likelihood Ratio</i>	6,021	sig 0,049
Tau de Goodman e Kruskal	0,03	sig 0,055
V de Cramer	0,27	sig 0,037

É difícil estabelecer o quanto estes resultados são confiáveis, tendo em vista que este grupo de fatores não se mostrou significativo para nenhuma das demais variantes. É possível que estes resultados sejam devidos à grande diferença entre as frequências observadas para a presença e a ausência do conector, e que a significância foi alcançada pelo fato das orações justapostas serem em número suficientemente grande para tal.

### 4.3 Resultados da amostra de fala de Florianópolis (VAR SUL)

A amostra de Florianópolis conta com apenas 45 ocorrências. Em virtude deste número total ser tão baixo, não foi possível obter resultados significativos para nenhuma das variáveis independentes testadas nesta amostra. Será realizada apenas uma descrição da distribuição das ocorrências das construções em estudo no falar de Florianópolis, comparando estas distribuições, quando possível, às obtidas nas amostras maiores (textos escritos e todas as capitais da região Sul).

Na tabela 31 é apresentada a distribuição das ocorrências da variável “configuração da sentença” de acordo com as características sociais dos informantes (grupos de fatores extralingüísticos). Neste recorte da amostra de fala foram coletadas 45 ocorrências as sentenças complexas em estudo, destas, 5 (11%) correspondem a casos de alçamento, 26 (58%) a casos de configuração canônica, 7 (15,5%) para a oração infinitiva anteposta à matriz e outros 7 (15,5%) para sentenças com sintagma nominal apagado.

No tocante ao grupo de fatores “sexo”, a amostra de Florianópolis seguiu o mesmo padrão obtido na amostra de fala das capitais da região Sul, apresentando uma distribuição bastante equilibrada para os dois sexos em todas as variantes analisadas.

Na variável “faixa etária”, os informantes do grupo mais jovem (15 a 24 anos) produziram a maior parte das ocorrências coletadas. Comparativamente, na amostra de fala total, foi a faixa etária dos 25 a 49 anos aquela em que se obteve o maior número de ocorrências. É possível imaginar que as construções em estudo são mais freqüentes quanto mais jovem for o falante, entretanto, no momento não há dados disponíveis da faixa etária dos 15 a 24 anos das demais capitais da região Sul, para que seja possível averiguar se esta tendência se manifesta de fato em todas as cidades consideradas.

As mesmas observações valem se forem consideradas apenas as construções de alçamento a sujeito, este fenômeno parece ser mais recorrente quanto mais jovem é o falante, tanto no recorte de Florianópolis quanto em toda a amostra de fala.

Tabela 31

**Variável configuração da sentença na amostra de fala de Florianópolis – fatores sociais**

		CONFIGURAÇÃO DA SENTENÇA								Total	
		Alçamento		Ordem canônica		Oração inf. desl.		Elipse do SN			
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Sexo	feminino	3	60%	12	46,2%	4	42,9%	3	42,9%	22	48,9%
	masculino	2	40%	14	53,8%	3	57,1%	4	57,1%	23	51,1%
Faixa etária	entre 15 e 24 anos	4	80%	18	69,2%	3	42,9%	3	42,9%	28	62,2%
	entre 25 e 49 anos	1	20%	5	19,2%	3	42,9%	3	42,9%	12	26,7%
	mais de 50 anos	0	0%	3	11,5%	1	14,3%	1	14,3%	5	11,1%
Nível de escolaridade	1 a 4 anos	0	0%	8	30,8%	6	85,7%	2	28,6%	16	35,6%
	5 a 8 anos	4	80%	13	50%	0	0%	4	57,1%	21	46,7%
	9 a 11 anos	1	20%	5	19,2%	1	14,3%	1	14,3%	8	17,8%
Total		5	100%	26	100%	7	100%	7	100%	45	100%

Em relação ao grupo de fatores “nível de escolaridade”, na amostra de Florianópolis informantes com 1 a 4 anos de escolaridade não produziram nenhuma ocorrência de alçamento a sujeito, preferindo utilizar a forma canônica ou o deslocamento de toda a oração infinitiva. Já na amostra geral de fala, os informantes com 1 a 4 anos de escolaridade e aqueles com 5 a 8 anos foram os que mais utilizaram as construções de alçamento. Com base nos resultados obtidos na amostra de geral de fala, esperava-se a ocorrência de elipses em um número bem maior do que o de alçamentos (aproximadamente o dobro), entretanto a diferença observada no uso destas duas variantes na amostra de Florianópolis é bem pequena, já que há apenas duas ocorrências a mais de elipses do que de alçamentos.

Na tabela 32 são apresentadas as freqüências para a variável “configuração da sentença” de acordo com os grupos de fatores lingüísticos.

Os resultados para a variável dependente “configuração da sentença” na amostra de Florianópolis foram, concernente ao alçamento a sujeito, bem diferentes das outras amostras (capitais da região Sul e textos escritos de jornais). O número de construções alçadas foi bastante baixo, e a maioria ocorreu sem que houvesse nenhum conector entre as orações combinadas, ao contrário das outras duas amostras, em que a maior parte das construções de alçamento ocorreu concomitante à preposição *de*.



**Tabela 32**  
**Variável configuração da sentença na amostra de fala de Florianópolis – fatores lingüísticos**

		CONFIGURAÇÃO DA SENTENÇA									
		Alçamento		Ordem canônica		Oração infinitiva deslocada		Elipse do SN		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma de conexão entre as orações	ausência	4	80	18	69,2%	4	57,1%	3	42,9%	29	64,4%
	preposição <i>de</i>	1	20	6	23,1%	0	0%	3	42,9%	10	22,2%
	preposição <i>para</i>	0	0	2	7,7%	3	42,9%	1	14,3%	6	13,3%
Forma do SN (objeto do verbo infinito)	elipse	0	0%	0	0%	1	14,3%	7	100%	8	17,8%
	SN definido	1	20%	13	11,5%	2	28,6%	0	0%	6	13,3%
	SN genérico	4	80%	1	34,6%	2	28,6%	0	0%	15	33,3%
	SN indefinido	0	0%	26	50%	2	28,6%	0	0%	15	33,3%
	pronomes	0	0%	1	3,8%	0	0%	0	0%	1	2,2%
Exp. do sujeito do infinitivo	não expresso	3	60%	16	61,5%	7	100%	6	85,7%	32	71,1%
	expresso	2	40%	10	38,5%	0	0%	1	14,3%	13	28,9%
Presença de materiais interv.	ausente	4	80%	22	84,6%	6	85,7%	6	85,7%	38	84,4%
	presente	1	20%	4	15,4%	1	14,3%	1	14,3%	7	15,6%
Distância da última menção	primeira menção	0	0%	10	38,5%	1	14,3%	0	0%	11	24,4%
	1 a 3 orações	4	80%	12	46,2%	5	71,4%	7	100%	28	62,2%
	+ de 4 orações	1	20%	4	15,3%	1	14,3%	0	0%	6	13,4%
Persistência nas 10 orações subsequentes	sem retomadas	0	0%	12	46,2%	4	57,1%	3	42,9%	19	42,2%
	1 a 2 retomadas	3	60%	8	30,8%	3	42,9%	2	28,6%	16	35,6%
	3 ou + retomadas	2	40%	6	23%	0	0%	2	28,6%	8	17,8%
Status informacional	dado	5	100%	16	61,5%	67	85,7%	7	100%	34	75,6%
	inferível	0	0%	2	7,7%	0	0%	0	0%	2	4,4%
	novo	0	0%	8	30,8%	1	14,3%	0	0%	9	20%
Total		5	100%	26	100%	7	100%	7	100%	45	100%

Em relação à forma do SN, na amostra de Florianópolis os tipos mais alçados foram os sintagmas nominais genéricos. Na configuração canônica, os sintagmas nominais indefinidos foram os mais utilizados. Tanto na amostra total do VARSUL quanto na do NILC, os sintagmas mais alçados foram os definidos, sendo que estes também foram as formas de SN mais comuns na ordenação canônica.

Quanto à expressão do sujeito do verbo infinitivo, os resultados foram semelhantes em Florianópolis e nas demais amostras, com sujeitos não expressos dominando a maioria das ocorrências, e a configuração em que mais foi utilizado o sujeito do infinitivo foi a ordem canônica.

Observando a distribuição das ocorrências no grupo de fatores “distância referencial” e “status informacional”, nota-se que estas construções tipicamente não tem a função de recuperar referentes muito distantes, tampouco introduzir referentes novos no discurso falado. No grupo “distância referencial”, tanto na amostra de Florianópolis quanto na das três capitais juntas, o valor modal ficou na categoria “1 a 3 orações” para qualquer

configuração da sentença considerada, o mesmo ocorrendo com o grupo de fatores “status informacional” na categoria “dado”.

Na amostra de textos escritos, a maioria dos alçamentos também ocorreu quando o referente (item alçado) estava a uma distância entre 1 e 3 orações da sua última menção. Entretanto, quando a configuração utilizada foi a ordem canônica, a maioria dos referentes estava sendo mencionado pela primeira vez. Isso aliado ao fato de que a maioria das ocorrências de textos escritos tem o status informacional “novo” ou “inferível”, indica que, na escrita, as sentenças do tipo estudado aqui também podem exercer a função de introduzir novos referentes no discurso. Na amostras de língua falada, ao contrário, a maioria dos referentes já era um informação dada no discurso precedente. É importante lembrar que o gênero textual da amostra falada é entrevista, então algumas vezes o referente em questão já havia sido mencionado pelo entrevistador.

Finalizando a descrição da amostra de Florianópolis, são apresentados na tabela 33 os resultados dos cruzamentos entre a variável dependente “forma de conexão entre as orações” e os grupos de fatores “item qualificador”, “expressão do sujeito do infinitivo”, “sexo”, “faixa etária”, “nível de escolaridade”.

**Tabela 33**  
**Variável forma de conexão entre as orações na amostra de fala de Florianópolis**

		FORMA DE CONEXÃO ENTRE AS ORAÇÕES							
		Ausência de conector		Preposição <i>de</i>		Preposição <i>para</i>		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Forma do elemento nominal	formas adjetivas	25	86,2%	8	80%	5	83,3%	38	84,4%
	formas substantivas	4	13,8%	2	20%	1	16,7%	7	15,6%
Exp. do sujeito do infinitivo	não expresso	19	65,5%	9	90%	4	66,7%	32	71,1%
	expresso	10	34,5%	1	10%	2	33,3%	13	28,9%
Sexo	feminino	15	51,7%	5	10%	2	33,3%	22	48,9%
	masculino	14	48,3%	5	10%	4	66,7%	23	51,1%
Faixa etária	entre 15 e 24 anos	21	72,4%	5	50%	2	33,3%	28	62,2%
	entre 25 e 49	7	24,1%	3	30%	2	33,3%	12	26,7%
	mais de 50	1	3,4%	2	20%	2	33,3%	5	11,1%
Nível de escolaridade	de 1-4 anos	8	27,6%	4	40%	4	66,7%	16	35,6%
	de 5-8 anos	16	55,2%	5	50%	0	0%	21	46,7%
	de 9-11 anos	5	17,2%	1	10%	2	33,3%	8	17,8%
Total		29	100%	10	100%	6	100%	45	100%

A distribuição geral das ocorrências na variável dependente “forma de conexão entre as orações” divide-se em 29 ocorrências de justaposição (64,4%), 10 ocorrências da preposição *de* (22,2%) e apenas 6 ocorrências da preposição *para* (13,3%).

A maioria das ocorrências (71,1%) não apresentou sujeito do infinitivo expresso. Este dado, aliado ao fato de que a maioria absoluta das ocorrências foi realizada sem que houvesse conector entre as orações encaixadas, faz com que a variação observada entre a expressão ou não do sujeito seja maior quando não há conector presente na sentença.

Quando a forma de conexão é a preposição *de*, em nove das dez ocorrências deste conector não houve realização do sujeito do infinitivo. De modo geral, na fala de Florianópolis não foi registrado um uso muito comum da preposição *para* nos contextos controlados.

Comparando a recorrência das diferentes formas de conexão com os demais resultados obtidos, observa-se este mesmo padrão: na maioria das ocorrências é realizada a justaposição das orações encaixadas, as ocorrências com o conector *de* vêm em segundo lugar e, por último, as ocorrências de *para*.

Neste recorte da amostra de fala também não foi registrado um uso diferenciado dos conectores entre homens e mulheres. Em relação à faixa etária, os mais jovens foram os informantes que mais produziram a construção em estudo, sendo que as ocorrências se distribuem em uma escala decrescente em todas as variáveis, à exceção da preposição *para*, que foi realizada em igual número de vezes por todas as faixas de idade.

Considerando o nível de escolaridade, é interessante notar que o nível intermediário (5-8 anos) não apresentou nenhuma ocorrência de *para*, entretanto, há poucas ocorrências no geral, especialmente de *para*. Esta faixa etária também foi a que mostrou a preferência mais clara pela conexão por justaposição.

## 5 CONSIDERAÇÕES FINAIS

O foco principal da pesquisa recai sobre as construções de alçamento, procurando descobrir se existe uma relação entre este fenômeno e alguma das formas de conexão das orações desta sentença complexa, entre outras características da estrutura. Uma das idéias norteadoras da pesquisa se referia à possibilidade do alçamento a sujeito estar relacionado à organização mais geral de uma fatia de discurso multi-proposicional. Considerando a “configuração da sentença” como variável dependente, o levantamento de dados e as análises realizadas apontam, tanto na fala quanto na escrita, para uma forte relação entre a configuração e a forma de conexão entre as orações combinadas.

Analizando a variação no uso das três possibilidades de conexão entre as orações combinadas em estudo, quais sejam, a ausência de conector (justaposição das orações); preposição *de*; e preposição *para*, observou-se que, cada forma de conexão é utilizada em um contexto lingüístico mais característico.

A justaposição na fala está relacionada ao uso de sintagmas definidos ou indefinidos como objetos do verbo no infinitivo, bem como à ausência de materiais intervenientes na sentença. É importante lembrar que poucas sentenças apresentaram materiais intervenientes nesta amostra. A justaposição das orações encaixadas é o ambiente que permite a maior liberdade de ordenação dos elementos da sentença. Na justaposição a oração reduzida pode permanecer na ordem canônica ou ser inteiramente deslocada à esquerda, como também permite o alçamento do objeto da oração encaixada para a posição de sujeito da oração matriz.

Tanto na escrita quanto na fala, o emprego da preposição *de* está fortemente relacionado à realização do objeto do infinitivo na forma de elipse (anáfora zero), ou de pronome. Na escrita, a preferência é clara para os pronomes do tipo resumitivo/fórico, já na fala, pronomes pessoais e resumitivos/fóricos tiveram uma distribuição bastante similar

Na fala, o fenômeno do alçamento a sujeito é bastante influenciado pelo uso da

preposição *de* como conector inter-oracional. Na, escrita o alçamento a sujeito também está relacionado ao uso da preposição *de*, e aliado à forma de conexão. Uma explicação para a influência do uso da preposição *de* sobre o alçamento envolve a reanálise da estrutura sintática na ordem canônica da sentença. A configuração canônica esta relacionada à conexão entre as orações por justaposição, tanto na fala ou na escrita. Observe-se o exemplos:

(64) Mas a bicicleta, naquele tempo, era difícil adquirir a bicicleta nova (VARSL FLP 04).

(64a) [era difícil [adquirir [a bicicleta nova]]].

A estrutura proposta em (64a) seria a interpretação mais tradicional da oração reduzida, na qual a sua classificação seria a de subjetiva (sujeito), da oração matriz. Neste caso, tem-se uma oração subordinada que é um constituinte da oração matriz, em uma situação que preenche diversos requisitos de encaixamento (cláusula reduzida, perda do sujeito, perda de liberdade de ordenação dos constituintes internos). O levantamento de dados realizado nesta pesquisa trouxe evidências que outra análise pode ocorrer, quando a oração é deslocada para a esquerda ou quando o objeto do sujeito do infinitivo é alçado a uma posição mais tópica.

(65) Mas, às vezes, dispensar uma cerveja fica difícil (NILC, es94de04).

(66) Depende, a tramitação no Congresso é difícil prever (NILC, ci96mr26).

A opção pelo deslocamento da oração combinada ou pelo alçamento não são determinações estruturais, tanto que estas formas são marcadas no que se refere à frequência de ocorrência. Neste caso, o alçamento do objeto do verbo infinitivo à posição de sujeito é uma opção organizacional do discurso feita pelo usuário da língua. Esta variabilidade posicional levanta um argumento contrário ao encaixamento, por denunciar uma ligação mais fraca entre as orações, a relação entre elas seria então de hipotaxe. Outra base de apoio para a interpretação contrária ao encaixamento nestas construções está na evidência de que a configuração preferencial da sentença complexa é a ordem canônica quando o sujeito do infinitivo é expresso, e neste caso, a dessentencialização da oração combinada é ainda menor.

Construções de alçamento, porém, praticamente não aceitam a expressão de sujeito do infinitivo. Além disso, o verbo no infinitivo tem a chance de se tornar ainda mais nominalizado, graças à ausência de seu complemento (que também ocorre quando há o apagamento do complemento deste verbo). Nos casos de alçamento, o uso do conector *de*

precedendo o infinitivo pode ser explicado pela reanálise do verbo nominalizado como um complemento nominal, tornando o uso desta preposição praticamente obrigatório nestes contextos.

(67) Além disso, o processo de microimpressão é difícil de reproduzir (NILC, ec96mr26)

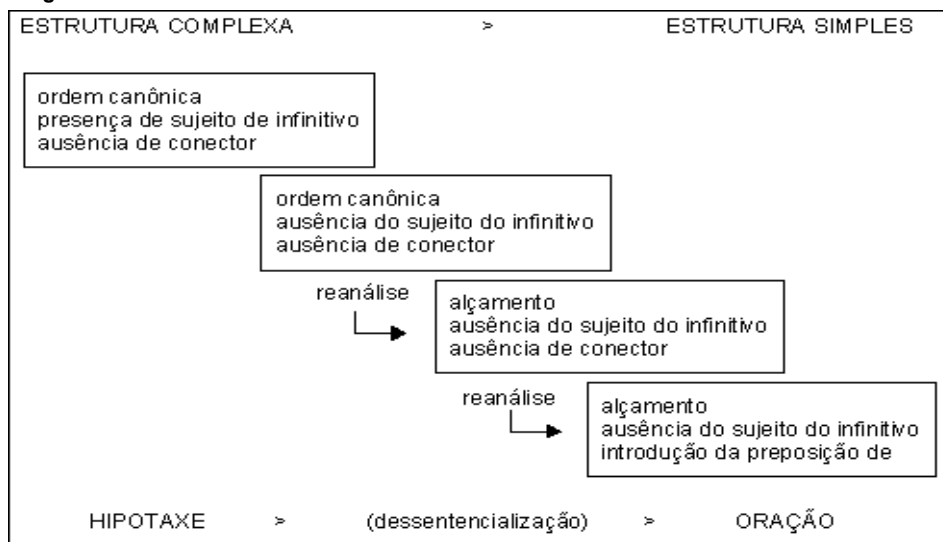
(68) Tamanha pressão não é fácil de suportar. (NILC, br94ab08).

(69) A tecnologia Aple é reconhecidamente superior, mais fácil de usar (NILC, ci96fe27).

Neste ponto, a estrutura se encontra entrelaçada a tal ponto que pode ser compreendida como uma sentença simples, não mais complexa. Outra evidência de que as estruturas combinadas através do conector *de* são mais integradas está na aparente impossibilidade de ocorrência destas sentenças antepostas à matriz. Tanto na fala quanto na escrita, não foram encontradas ocorrências de oração infinitiva deslocada introduzida pela preposição *de*. Isto sugere que *de*, ao contrário de *para*, não exerce a função de conjunção inter-oracional.

O alto nível de dessentencialização e de entrelaçamento apresentado pelas construções de alçamento com a preposição *de* sugere também que esta construção é mais gramaticalizada do que as construções combinadas por justaposição. Neste caso, a preposição *de* não é um conector inter-oracional, mas uma preposição gramatical cuja função é atribuição de caso ao sintagma nominal (verbo nominalizado). Não há aqui nenhuma violação aos parâmetros propostos por Lehmann (1988). O percurso desta estrutura num continuum de gramaticalização poderia ser esquematizado da seguinte forma ilustrada no diagrama 1.

Diagrama 1



As caixas empilhadas no diagrama representam as diversas camadas de organização gramatical das sentenças. No momento atual, todas as estão em uso, sendo que a estrutura na ordem canônica com ausência de sujeito do infinitivo é a forma mais utilizada. Não espera-se que ocorra o desaparecimento das demais formas, mas sim que haja um aumento significativo na frequência de uso das formas mais integradas.

Utilizando o mesmo tipo de raciocínio pode-se compreender a introdução do conector *para* nas construções estudadas. O uso do conector *para* é em grande parte condicionado, tanto na fala quanto na escrita, pela expressão do sujeito do verbo infinitivo. Neste caso, temos duas formas variantes disputando para a mesma função, uma sentença combinada por justaposição e uma sentença combinada através de *para*. Exemplos:

(56) Antes de quinze dias era difícil você conseguir um ortopedista. (VARSUL, CTB 19)

(57) Então eu acho assim que se não houver a comunicação, o diálogo, né? eu acho que daí é mais difícil pra ti superar o casamento (VARSUL, POA 12).

As sentenças combinadas introduzidas por *para* também tem uma relativa liberdade posicional, já que podem ocorrer tanto antepostas quanto pospostas em relação à matriz. Na escrita não foram encontradas muitas ocorrências nas quais a oração infinitiva inteira foi anteposta à matriz. Na fala, esta configuração ocorreu tanto quanto as construções de alçamento. Os fatores mais relevantes para a realização do deslocamento de toda a oração infinitiva são a presença da preposição *para* e a expressão do sujeito do verbo finito. Isto traz evidências de que esta forma de conector é utilizada para introduzir um constituinte oracional (diferentemente da preposição *de*).

As orações introduzidas pela preposição *para* apresentam mais restrições quanto à organização interna dos seus elementos, já que a reduzida não é “despedaçada” quando este conector é utilizado, ou seja, construções em que o objeto do verbo infinitivo é alçado não são contextos típicos do uso de *para*. Orações combinadas por *para*, por conseguinte, podem ser identificadas como estruturas hipotáticas, se forem considerados os parâmetros propostos por Lehmann (1988). Assim, a estrutura combinada com *para* é menos integrada do que aquela que apresenta a preposição *de*.

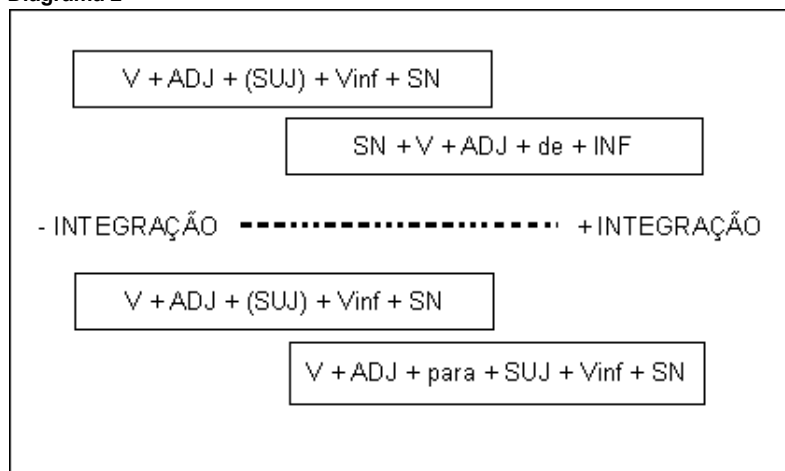
Em relação às estruturas justapostas, é mais difícil identificar qual a estrutura mais integrada, a que apresenta ou a que não apresenta conector. Por um lado, as orações combinadas pelo conector *para* apresentam maiores restrições quanto à ordenação interna de

seus constituintes, já que não são contextos preferenciais de alçamento. Por outro lado, a ausência de conector é um dos parâmetros de maior integração entre orações combinadas. Entretanto, as construções combinadas com *para* não apresentam a ambigüidade estrutural apresentada pelas orações justapostas.

É possível especular que a introdução do conector em tais construções esteja ocorrendo para eliminar a ambigüidade estrutural apresentada pelas estruturas justapostas. Espera-se então, que o uso das construções com conector aumente cada vez mais. Isso conduziria à situação apresentada no diagrama 2.

Nesta situação, as estruturas com conector não variam entre si, mas sim com as estruturas justapostas. Entretanto, para que esta situação se configure, é necessário um aumento significativo no uso das construções com a presença do conector. Além disso, é preciso considerar a atuação de outros fatores, que podem modificar completamente este quadro.

**Diagrama 2**



Ainda que as variáveis discursivas não tenham se mostrado tão influentes na ordenação dos constituintes quanto esperado inicialmente, não parece apropriado descartar a atuação de influências discursivas e pragmáticas sobre a configuração destes tipos de sentença. Em primeiro lugar, as categorias discursivas analisadas neste estudo não são as únicas possíveis, em segundo lugar, é provável que a topicalidade possa ser melhor descrita em uma única variável independente complexa, que represente diferentes graus desta propriedade, do que através de duas ou três variáveis discretas. Além disso, estratégias de organização textual podem ainda ser investigadas, relacionadas a propriedades tais como o gênero do discurso e coesão.



Resultados mais iluminadores podem ser atingidos também com o aumento da amostra de língua falada, que se mostrou o ambiente mais rico em termos de variação no uso das diferentes formas das construções aqui estudadas. Outra perspectiva também é a realização de estudo similar no espaço de tempo de alguns anos, de modo a verificar se ocorreu aumento no uso das construções com conector, hipótese sustentada aqui, ou se, pelo contrário, houve a estabilização ou recrudescimento no uso destas formas. Neste último caso, seria interessante avaliar o que provocou a diminuição, e contrastar os resultados obtidos então com aqueles expostos neste trabalho.

Espera-se, assim, que tenha sido possível oferecer um trabalho em que outros possam tomar como ponto de partida, para que se possa, sobretudo, enriquecer e melhorar o conhecimento acumulado até o momento.

## REFERÊNCIAS

- ALI, S. M. (1965) **Gramática secundária da língua portuguesa**. São Paulo: Melhoramentos.
- ALI, S. M. (1971) **Gramática histórica da língua portuguesa**. Rio de Janeiro: Livraria Acadêmica.
- ALMEIDA, N. M. (1962) **Gramática metódica da língua portuguesa**. São Paulo: Saraiva.
- ALUÍSIO, S.M.; PINHEIRO, G.M.; FINGER, M.; NUNES, M.G.V.; TAGNIN, S.E. (2003) The Lacio-Web Project: overview and issues in Brazilian Portuguese corpora creation. In: ARCHER, D.; RAYSON, P.; WILSON, A.; MCENERY, T. (eds.). **Proceedings of the Corpus Linguistics**, v. 16, part 1, p. 14-21.
- BECHARA, E. (1976) **Moderna gramática da língua portuguesa**. São Paulo: Nacional.
- BECKER, L. A. (1999) **CROSSTABS**: Measures for Nominal Data. Disponível em: <<http://web.uccs.edu/lbecker/SPSS/ctabs1.hhtm>>.
- BYBEE, J. (2003) Cognitive processes in grammaticalization. In.: TOMASELLO, M. (ed.) **The new psychology of language**. New Jersey: Lawrence Erlbaum Associates, v. 2.
- CUNHA, C. F. (1977) **Gramática da língua português**. Rio de Janeiro: Fename.
- CUNHA, C. F.; CINTRA, L (1985) **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira.
- GIVÓN, T. (1983) The 'topic' strand: micro traditions. In: GIVÓN, T (ed.) **Topic continuity in discourse**: a quantitative cross-language study. Philadelphia: John Benjamins, p. 5-41.
- GIVÓN, T.. (1991) Isomorphism in the grammatical code: cognitive and biological considerations. In: **Studies in language**. Philadelphia: J. Benjamins, n. 15.
- GIVÓN, T. (1995) **Functionalism and Grammar**. Philadelphia: J. Benjamins.
- GIVÓN, T. (2001a) **Syntax**. Philadelphia: J. Benjamins, v. 1.
- GIVÓN, T.. (2001b) **Syntax**: an introduction. Philadelphia: John Benjamins, v. 2.
- GORSKI, E.; MITTMANN, M. M.; MAY, G. (2004) Um estudo pancrônico de orações reduzidas: considerações sobre conexão e parâmetros de integração. In.: **Anais da XX Jornada Nacional de Estudos Lingüísticos do GELNE**. João pessoa, UFPB, p. 671-680.
- GORSKI, E.; MITTMANN, M. M.; NEVES, P. (2003) O estatuto sintático camaleônico de

certas construções infinitivas: uma abordagem pancrônica. **Working papers em lingüística**. Florianópolis: UFSC, n. 7, p. 73-94.

HEINE, B.; CLAUDI, U.; HÜNNEMEYER, F. (1991) **Grammaticalization**: a conceptual framework. Chicago: University of Chicago.

HOPPER, P. (1987) Emergent grammar. **Berkeley Linguistic Society**, n. 13, p. 139-157.

HOPPER, P.; TRAUGOTT, E. (1993) **Grammaticalization**. Cambridge: Cambridge University.

KNIES, C. B.; COSTA, I. B. (orgs.) (1995) Manual do usuário: banco de dados lingüísticos "VARSQL". UFPR/UFSC/UFRGS/PUC-RS.

LABOV, W. (1978) (1978) Where does the linguistic variable stop? A response to Beatriz Lavandera. **Sociolinguistic Working Paper**, n. 44, p. 43-88.

LACIO-WEB (2003). **Lacio-Web**. Núcleo Interinstitucional de Lingüística Computacional. Disponível em: <<http://www.nilc.icmc.usp.br/lacioweb>>.

LEHMANN, C. (1988) Towards a typology of clause linkage. In.: HAIMAN, J.; THOMPSON, S. (eds.) **Clause combining in grammar and discourse**. Philadelphia: J. Benjamins.

MARTELOTTA, M. E.; VOTRE, S. J.; CEZARIO, M. M. (orgs.) (1996) **Gramaticalização no português do Brasil**: uma abordagem funcional. Rio de Janeiro: Tempo Brasileiro.

MAY, G. (2004) **"Isso é fácil de dizer", mas difícil de analisar**: construções subjetivas em perspectiva DIACRÔNICA. (Relatório de Pesquisa PIBIC/CNPq/UFSC).

MITTMANN, M. M. (2003) **De Gil Vicente a Miguel Falabella**: uma análise diacrônica de construções subjetivas. (Relatório de Pesquisa PIBIC/CNPq/UFSC).

MELO, G. C. de. (1978) **Gramática fundamental da língua portuguesa**. Rio de Janeiro: Ao Livro Técnico.

MOLLICA, M. C. (org.) (1992) **Introdução à sociolingüística variacionista**. Rio de Janeiro: UFRJ. (Cadernos didáticos UFRJ).

NEVES, P. (2003) **Uma análise sincrônica de construções subjetivas com verbo ser na fala de Florianópolis**. (Relatório de Pesquisa PIBIC/CNPq/UFSC).

NEVES, P. (2004) **Sociolingüística está legal (pra, de) estudar**: construções subjetivas analisadas sincronicamente. (Relatório de Pesquisa PIBIC/CNPq).

OITICICA, J. (1940) **Manual de estilo**. Rio de Janeiro: Francisco Alves.

PINHEIRO, G. M.; ALUÍSIO, S. M. (2003) **Corpus Nilc: descrição e análise crítica com vistas ao projeto Lacio-Web**. Série de Relatórios do NILC. Disponível em: <<http://www.nilc.icmc.usp.br/nilc/>>.

PINTZUK, S. (1988) **VARBRUL Program**. Philadelphia: University of Pennsylvania. 1988.

RAND, D.; SANKOFF, D. (1990) **GoldVarb**: a variable rule application for Macintosh.

ROBINSON, J.; LAWRENCE, H.; TAGLIAMONTE, S. (2001) **GOLDVARB 2001**: a

multivariate analysis application for windows. Disponível em:

<<http://www.york.ac.uk/depts/lang/webstuff/goldvarb/manual/manualOct2001.html>>.

ROCHA, M. A de F. (1989). Complementizadores no português do Brasil: uma abordagem inter- e intra-sistêmica. In: TARALLO, F. **Fotografias sociolingüísticas**. Campinas: Universidade Estadual de Campinas, p. 141-163.

ROCHA LIMA, C. H. (1986) **Gramática normativa da Língua Portuguesa**. Rio de Janeiro: José Olympio.

ROCHA LIMA, C. H. (1972) **Gramática normativa da língua portuguesa**. Rio de Janeiro: José Olympio.

SEAMAN, M. A. (2001) **SPSS: Categorical Data**. Disponível em:

<<http://edpsych.ed.sc.edu/seaman/edrm711/questions/categorical.htm>>.

SARDINHA, T. B. (2004) **Lingüística de corpus**. Barueri: Manole.

SCOTT, M. (1999) **WordSmith Tools 3.00.0**. Oxford University Press.

SPSS (1999) **SPSS for Windows 10.0.1**. SPSS Inc.

VOTRE, S. (1992) **Lingüística funcional: teoria e prática**. Rio de Janeiro: UFRJ.

WEINREICH, U.; LABOV, W.; HERZOG, M. (1978) **Empirical foundations for a theory of language change**. Austin: University of Texas.